



Dynamique de la Motorisation et Usage de l'Automobile en France (L'Île-de-France en Perspective)

Roger Collet

► To cite this version:

Roger Collet. Dynamique de la Motorisation et Usage de l'Automobile en France (L'Île-de-France en Perspective). Economies et finances. Université Panthéon-Sorbonne - Paris I, 2007. Français. NNT : . tel-00318707

HAL Id: tel-00318707

<https://theses.hal.science/tel-00318707>

Submitted on 4 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamique de la Motorisation et Usage de l'Automobile en France *L'Île-de-France en Perspective*

THÈSE pour le doctorat en Sciences Économiques
présentée et soutenue publiquement par

ROGER COLLET

le 13 novembre 2007

Directeur de thèse :

Professeur François GARDES, Université de Paris I – Panthéon-Sorbonne

Composition du jury :

Professeur Georges BRESSON, Université de Paris II – Panthéon-Assas
(rapporteur)

Docteur Laurent HIVERT, Ingénieur Divisionnaire des Travaux Publics de
l'État, INRETS-DEST

Docteur Jean-Loup MADRE, Directeur de Recherche, INRETS-DEST

Professeur Philip MERRIGAN, Université du Québec à Montréal
(rapporteur)

Professeur Christian ROBERT, Université de Paris-Dauphine

Professeur Jean-Marc ROBIN, Université de Paris I – Panthéon-Sorbonne
(président du jury)

L'UNIVERSITE DE PARIS I PANTHEON-SORBONNE n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

A mes arrière-grands-parents,
Thérèse, Jacqueline, et Gilbert.

Remerciements

Libéré de la rigueur de la thèse, l'écriture des remerciements devrait être une simple formalité, où, passant en revue tous ceux qui ont apporté leur bon concours, le cœur suffirait à donner la bonne mesure de la reconnaissance envers chacun. Pourtant, il s'agit d'un exercice de style très particulier, dont l'enjeu serait davantage l'expression d'une profonde sincérité que de simplement remercier, et dont le paradoxe serait de se contenir et rester solennel, alors que l'euphorie d'avoir terminé la rédaction est débordante. C'est donc sans retenue que je me lance dans cet exercice, en abusant des superlatifs, relatifs et absolus, et des déclinaisons du mot « merci ».

Mes premiers remerciements, très chaleureux, sont adressés au Professeur François Gardes, de l'université de Paris I. En acceptant de devenir mon directeur de recherche, il me fit le grand honneur de sa confiance, et le plaisir d'un suivi à la fois attentif et enrichissant. Sa faculté à remobiliser ma volonté lorsque mes recherches piétinaient et la grande disponibilité qu'il m'a toujours témoignée m'auront beaucoup aidé dans ma rédaction. Je lui suis très reconnaissant d'avoir ouvert son carnet d'adresse en 2002, pour me trouver dans ses relations une opportunité d'effectuer une thèse financée dans un pôle de recherche.

C'est grâce à sa bienveillance que j'ai pu rencontrer Jean-Loup Madre, directeur du Département d'Economie et de Sociologie des Transports (DEST), à l'Institut National de REcherche sur les Transports et leur Sécurité (INRETS), auprès de qui je suis également très redevable. Je lui adresse mes plus vifs remerciements pour m'avoir aidé à monter le projet de thèse, pour son accueil au DEST en 2003, et pour avoir mis à ma disposition des conditions de travail idéales jusqu'à mon départ à la mi 2006.

Francis Papon, puis Laurent Hivert, chercheurs à l'INRETS-DEST, m'ont fait le privilège d'être les encadrants de mes travaux de recherche, je les remercie d'avoir pris cette responsabilité. J'ai profité de leurs connaissances et de leur précieuse aide pour investir les données traitées dans ce mémoire. Avec Jean-Loup Madre, leur clairvoyance des problématiques automobiles m'a paru très impressionnante. Plus particulièrement, je souhaite remercier Laurent Hivert pour son précieux effort de relecture, marqué par de nombreux ajouts, remarques et corrections. A l'INRETS, je remercie également tous les curieux qui se sont penchés avec intérêt sur mes travaux et qui m'ont apporté astuces et commentaires très fructueux, notamment les trois mousquetaires-économètres : Akli Berri, Matthieu De Lapparent, et Diègo Legros.

Je suis également reconnaissant envers Christophe Starzec, chercheur au CNRS et à l'université de Paris I, pour m'avoir fait partager sa compréhension très éclairante des modèles d'addiction, dont j'ai fait une utilisation au cours de cette recherche.

Malgré toute l'aide reçue, je reste bien entendu l'unique responsable des erreurs ou des insuffisances qui, éventuellement, subsisteraient dans ce document.

Pour tout l'honneur qu'ils me font d'accorder leur grande expertise à mon travail et de l'évaluer, j'exprime ma gratitude la plus profonde aux membres de mon jury de thèse : Georges Bresson (Université Panthéon-Assas, Paris II), François Gardes (Université Panthéon-Sorbonne, Paris I), Laurent Hivert (INRETS-DEST), Jean-Loup Madre (INRETS-DEST), Philip Merrigan (Université du Québec, Montréal), Christian Robert (Université de Paris-Dauphine), Jean-Marc Robin (Université Panthéon-Sorbonne, Paris I).

Pour leur soutien financier, j'exprime ma gratitude envers l'INRETS et le conseil régional d'Île-de-France qui ont co-alimenté mon allocation de recherche de 2003 à 2005.

Dans un répertoire plus intime, je remercie mon entourage familial proche d'avoir été compréhensif, malgré ma maladresse à lui faire saisir toute la difficulté de mon projet. Avec son achèvement, je suis heureux de le libérer de ses inquiétudes, particulièrement Brigitte, ma mère. Merci aussi à mon frère, Frédéric, dont le courage exemplaire à travailler tôt chaque matin (ou tard chaque soir) m'aura servi d'exemple pour ne pas (trop) me relâcher, lorsque je suis revenu à la maison afin de terminer ma rédaction. J'ai également une pensée pour Anne-Marie, Georges, Michel (a), Franck et sa nouvelle petite famille, ainsi qu'une toute particulière pour Michel (b). Roger, mon père, je ne t'oublie bien évidemment pas.

Je salue aussi les quelques âmes délicates qui, lorsque nous nous sommes vus, se sont toujours inquiétées de l'évolution de mes projets, et qui acquiesçaient poliment lorsque je rentrais un peu dans le détail de mes travaux : David, Inès, Natacha, Coralie, Patrick, et Julien.

Bien sûr, je tiens à remercier du plus profond de mon cœur mon amie, Tatiana, dont l'amour m'a toujours été très précieux, particulièrement lorsque mes états d'âme de thésard se faisaient sombres et orageux. Son écoute, sa compréhension, et son soutien m'ont été d'un réconfort indispensable à retrouver ma sérénité. Merci à toi, Tatiana, d'avoir eu la patience de m'attendre, et d'avoir toujours su trouver les mots pour m'encourager dans mon projet. Ces années durant, ton sourire m'a été une source inépuisable d'énergie, renouvelée au quotidien.

Enfin, pour leur soutien, leur aide et leur confiance infaillibles depuis toujours et jusqu'à jamais, je dédie mon travail à mes arrière-grands-parents. A Thérèse, et à la mémoire de Jacqueline et Gilbert, qui m'ont absolument tout donné. Je leur suis éternellement reconnaissant.

Sommaire

<i>REMERCIEMENTS</i>	7
SOMMAIRE	9
INTRODUCTION GENERALE	11
CHAPITRE 1 : LES DONNEES DE L'ENQUETE «PARC AUTOMOBILE».....	25
CHAPITRE 2 : ADDICTION A L'USAGE DE L'AUTOMOBILE ET MESURES D'ELASTICITES	37
CHAPITRE 3 : DYNAMIQUE DU NIVEAU DE MOTORISATION DES MENAGES	111
CHAPITRE 4 : ETUDE DU CHOIX D'ACQUISITION AUTOMOBILE	175
CONCLUSION GENERALE.....	235
REFERENCES BIBLIOGRAPHIQUES.....	247
ANNEXE : FORMULAIRE VOITURE DE L'ENQUETE « PARC AUTO » 2001	259
TABLE DES MATIERES	275
LISTE DES FIGURES	278
LISTE DES TABLES	278

Introduction générale

Introduction générale

Au sortir de la seconde guerre mondiale, la France, comme ses voisins en Europe de l'ouest (Royaume-Uni, Italie, ex-RFA...) a fait le choix du plan Marshall et de l'automobile. Les programmes d'investissement publics dans les infrastructures routières ont contribué à relancer l'activité économique, et ont favorisé la demande de voitures. En produisant à grande échelle des véhicules à bas prix, facilement écoulés sur un marché alors très demandeur, l'industrie automobile amorçait l'ère de la voiture populaire, et servait de tremplin à la reconstruction économique, en engageant notamment autour d'elle les industries métallurgiques, verrières, pneumatiques, pétrochimiques, les activités de crédit, ainsi que les services d'après vente (distribution du carburant, garages, assurances...). En contractant les espaces, en réduisant le temps, et en accroissant l'accessibilité, elle incarnait un nouveau mode de vie ouvert sur l'extérieur, la découverte, la communication, et devenait un emblème de la renaissance d'après guerre. Tous ses avantages avaient rapidement donné à l'automobile une dimension idéologique forte. Erigée en symbole de liberté et du capitalisme¹, elle s'opposait alors à l'encadrement des transports publics dans le bloc communiste des pays de l'est. En ces temps, bénie était l'automobile...

Mais le revers de la médaille ne tardait pas à trouver un écho. Dans les années 1960, quelques voix s'élèvent pour dénoncer le processus de motorisation en France (SAUVY, 1968), alors que l'extension du réseau routier appelle sans cesse de nouveaux usagers et ne suffit pas à résorber un nouveau phénomène récurrent : la congestion. L'automobile commence alors à devenir un sujet de controverses. L'augmentation du trafic s'est accompagnée de celle du nombre d'accidentés de la route, et des premiers préjudices sanitaires et environnementaux. Au début des années 1970, l'écologie et la sécurité routière s'invitent ainsi au débat, avec des arguments qui militent en faveur de l'action politique pour « civiliser » l'automobile. La France réalise avec le choc pétrolier de 1973 qu'elle n'a pas de pétrole², et mesure notamment sa dépendance à ses importations d'or noir pour mettre son parc en activité. A partir de cette date, les fluctuations autour de la tendance de croissance du marché automobile se sont amplifiées, et ont quelque peu

¹ Comme l'illustre cette aphorisme de Churchill : « Sous le capitalisme, les gens ont davantage de voitures. Sous le communisme, ils ont davantage de parkings ».

² « ...mais des idées », comme disait le slogan d'époque resté dans les mémoires.

ébranlé l'image de l'industrie automobile dans son rôle de moteur économique stable.

Les détracteurs de l'automobile lui font alors un procès qui remet en cause son sacre des trente glorieuses. Les points d'encombrement routier font perdre aux usagers une partie du gain de temps que l'automobile devait leur accorder, la croissance régulière du nombre de tués sur les routes inquiète, et met en cause des constructeurs responsables de vices de fabrication (NADER, 1968). Les signes de la pollution atmosphérique attribuée à l'automobile réveillent les préoccupations écologiques, et les exemples américains, comme Los Angeles enfumée sous le « smog » des gaz d'échappements, font déjà imaginer le pire. Enfin, la crainte de voir l'automobile absorber rapidement les ressources fossiles, ainsi que la facture énergétique alourdie par le second choc pétrolier de 1979 jettent l'incertitude sur la pérennité d'une société motorisée.

Dans un contexte d'hostilité grandissante, où l'automobile est cette fois décrite comme un symbole du gaspillage et d'aliénation à la société de consommation, ou comme le pilleur des ressources naturelles, l'ensemble des critiques aurait dû largement favoriser un report de mobilité sur les transports publics. Il n'en a rien été.

Tout d'abord parce que l'effet de relance induit par le contre-choc pétrolier du milieu des années 1980 et la baisse du prix du baril ont profité à l'usage automobile, et que les perspectives favorables de croissance économique ont relayé les tourments écologiques et énergétiques au second rang des priorités (GALLEZ, 1994).

Ensuite, le dynamisme de l'offre, dopé par l'approche *marketing*, a multiplié les gammes et les variétés de véhicules offerts sur le marché pour mieux répondre aux besoins individuels. Cette segmentation très fine de la clientèle s'est révélée une stratégie très efficace pour soutenir la demande automobile. Alors que déjà 7 ménages sur 10 sont équipés en 1984, et que le seuil de saturation semble s'établir autour de 80%, la progression du parc automobile prend corps dans la multi-motorisation des ménages. Anticipant d'après l'exemple nord-américain, la voiture familiale semble progressivement s'effacer au profit de la voiture personnelle (DUPUY, 1999 ; PAPON et HIVERT, 2006). Avec cette tendance à son individualisation, l'automobile devient le signe de réussite sociale, de dynamisme, de caractère, de virilité... Pour ses opposants, elle est dénoncée comme un signe

ostentatoire de richesse, de fatuité, de domination, d'agressivité, voire d'insuffisances...

Par ailleurs, la diffusion du moteur diesel (plus économe qu'un moteur essence « comparable ») dans toutes les variétés et à différents niveaux de gamme, accompagnée d'une fiscalité qui positionne le différentiel des prix des carburants à l'avantage du gazole, a réduit les coûts d'usage courant de l'automobile. Malgré un coût d'acquisition généralement plus élevé pour les voitures diesel, ce double aspect économique séduit encore aujourd'hui : c'est la « diésélisation » du parc automobile (HIVERT, 1996a, 1996b), dont la dynamique depuis le milieu des années 1980 a amplement changé sa physionomie.

Côté demande, les changements du profil de l'automobiliste français ont également appuyé le succès d'une offre automobile très diversifiée. L'accès progressif des femmes au permis de conduire depuis les années 1950 d'une part, et les générations de retraités dans les années 1990 qui ont connu l'essor de l'automobile et lui sont restées fidèles d'autre part, expliquent notamment la féminisation et la séniorisation des conducteurs, et constituent deux facteurs structurels ayant contribué à la vitalité du marché de l'automobile (HERPIN et VERGER, 1999).

L'action politique sur le marché de l'automobile reste un stimulant efficace pour soutenir son activité dans les périodes noires. Encore récemment, celle-ci servait de levier pour revigorer une économie en mal de croissance, avec des mesures gouvernementales subventionnant l'acquisition de véhicules neufs (les 'jupettes' en 1993, puis les 'balladurettes' en 1994³). Avec la crise de 1993, un enjeu était de dégager d'une spirale dépressionniste le grand ensemble du secteur automobile, dont dépendent un million cinq cent mille emplois au cours des années 1990 (CHANARON et LUNG, 1995)⁴.

En augmentant par ailleurs les distances entre les foyers et les bassins d'emploi, les zones d'activité, de soins, d'éducation, etc... la périurbanisation a eu en outre pour effet de renforcer les besoins en équipement automobile dans les foyers, créant ainsi une forme de dépendance.

La seconde moitié du 20^{ème} siècle est donc marquée par l'avènement de l'automobile en France. Aujourd'hui, routes, parkings, horodateurs, stations-service, centres de lavage, signalisations, passages pour piétons... nous rappellent que

³ Du nom des deux premiers ministres en poste à l'époque : A. Juppé, puis E. Balladur.

⁴ Cités dans HERPIN et VERGER (2000).

l'espace est résolument « automobilisé ». Malgré quelques voix protestataires, le plébiscite public pour la voiture a été tel que l'aménagement du territoire et l'urbanisme ont souvent fait de l'accès en automobile une priorité.

Dans l'usage, le succès de la voiture sur les autres modes de transports est sans conteste : en 1998, plus de huit kilomètres sur dix sont effectués en automobile (ORFEUIL, 2000) contre six dans les années 1950. Dans le même temps, les déplacements relativement courts caractérisent son utilisation : un sur deux est inférieur à cinq kilomètres, un sur quatre est inférieur à trois kilomètres (FONTAINE et HUBERT, 1998 ; ORFEUIL, 2000).

De cet historique, il apparaît que l'usage automobile, le degré de motorisation des ménages, et la diversité de qualité du parc constituent trois fondamentaux de l'équipement automobile. Dans ce travail, nous proposons d'en réévaluer les déterminants en France à l'aube du 21ème siècle, tout en mettant la région parisienne en perspective.

Relativement à sa population, l'Île-de-France est la région française la moins motorisée, mais également celle où les comportements automobiles sont les plus contrastés, notamment en comparant sa ville-centre (Paris intramuros), et sa grande couronne (MADRE, 1989). Dans un tissu urbain très dense, caractéristique de la région parisienne, l'usage intensif de l'automobile génère de fortes externalités négatives qui dégradent le bien-être des résidents et des utilisateurs de la route (augmentation de la pollution locale et des nuisances sonores, risque d'accident accru, phénomènes de congestion récurrents, retards et stress...voir la figure 0.1).

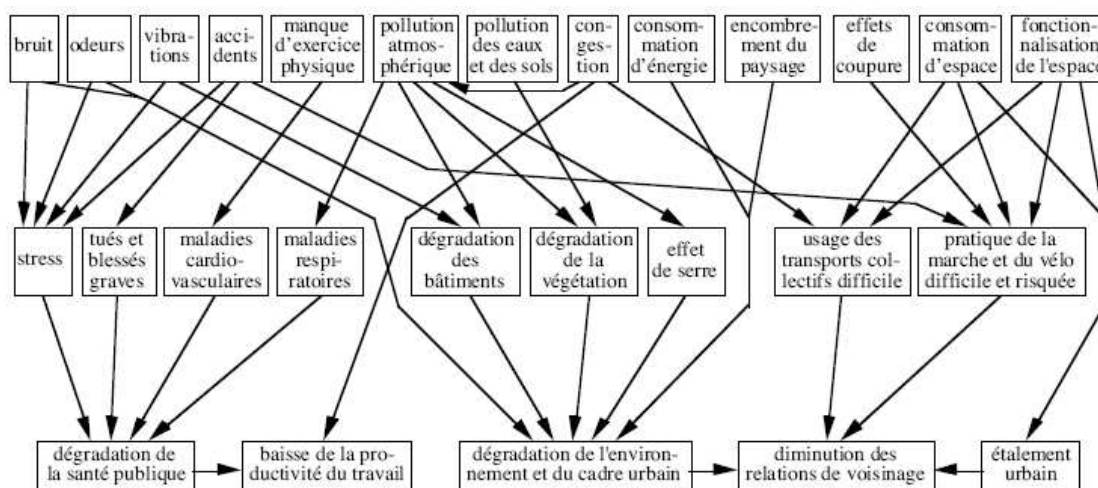
Au delà d'atténuer ces contraintes, localement acceptées par les résidents, travailleurs et automobilistes franciliens, l'enjeu de modérer l'usage automobile est devenu une priorité mondiale : l'avenir du climat. En effet, la France est tenue par ses accords internationaux, dont le protocole de Kyoto (1997), à la réduction de ses émissions de gaz à effet de serre⁵ auxquelles contribuent généreusement les activités de transport⁶, et donc, l'automobile⁷. Dans ce contexte, où les spécialistes

⁵ Dont la famille rassemble notamment : le dioxyde de carbone (CO₂), le méthane (CH₄), le protoxyde d'azote (N₂O), les perfluorocarbures (PFC), les hyperfluorocarbures (HFC), l'hexafluorure de soufre (SF₆). Par cet accord, la France s'est engagée à stabiliser ses émissions de 2010 à son niveau d'émissions en 1990.

⁶ 21% des émissions en France en 2004 (CITEPA, 2006 pour le Ministère de l'Environnement), et un secteur dont la croissance est restée l'une des plus élevée au cours du temps.

s'accordent pour dire que la technologie seule est actuellement insuffisante pour tout solutionner, l'urgence est à favoriser les changements de comportements, en allant vers une moindre pression de l'automobile dans les déplacements, et dissuader les agents qui tendent à un usage plus intensif des véhicules les moins « propres » et les plus « gourmands ».

FIGURE.0.1 : Les nuisances liées au trafic automobile en milieu urbain



NB : Ce schéma ne prend pas en compte les nombreuses boucles de rétroaction.

Source : HERAN (2001).

Nous marquerons donc un intérêt particulier à révéler les spécificités de la région Île-de-France, en confrontant les comportements automobiles des résidents franciliens avec ceux des Provinciaux. Mais aussi, nous comparerons les comportements intra Île-de-France, où le découpage zonal retenu correspond à Paris (département 75), la petite couronne francilienne (départements 92, 93, 94), et la grande couronne francilienne (départements 77, 78, 91, 95).

Nous choisissons de débuter notre recherche par l'étude de l'intensité d'usage par les ménages de leurs automobiles. Dans les faits, il s'agit du comportement le plus apparent et aux conséquences écologiques les plus immédiates. L'usage étant la

⁷ Avec les petits utilitaires, elle représente 15% des émissions en France en 2004 (CITEPA, 2006).

finalité de la décision d'équipement, nous remontons le processus décisionnel des ménages en analysant dans un second temps leur choix du niveau de motorisation. Nous affinons cette analyse quantitative avec l'étude qualitative des choix d'automobiles au niveau individuel dans une dernière étude.

Pour conduire nos analyses des trois pôles de la motorisation (usage, quantité, qualité), nous adoptons une approche désagrégée et statistique des phénomènes. Celle-ci nous a été rendue possible par les données de ménages et véhicules issues de l'enquête « Parc Automobile », disponibles à L'INRETS, dont nous avons utilisé les vagues annuelles de 1997 jusqu'en 2001. Une spécificité de l'enquête est la réinterrogation d'une partie des ménages au cours du temps, ce qui autorise le suivi temporel de leurs comportements de motorisation, et la disposition des données sous la forme d'un panel. Ce type de structure est un avantage certain lorsque l'on souhaite combiner simultanément l'analyse longitudinale et transversale, et pour asseoir des conclusions sur la dynamique comportementale. Dans notre plan, un tout premier chapitre, rapide, est donc consacré à la description de l'enquête « Parc Automobile », et à la construction d'un panel pondéré couvrant la période 1999-2001 (notamment utilisé dans le second et troisième chapitre).

Les protestations écologistes renaissantes du réchauffement climatique dénoncent aujourd'hui l'usage particulier de l'automobile comme l'origine la plus abusive d'émission des gaz à effets de serre⁸. Dans les grandes métropoles, comme à Paris, les pouvoirs publics se mobilisent pour limiter la circulation automobile et la pollution atmosphérique qu'elle génère, pour améliorer les conditions sanitaires des citoyens. Quelques élus militent même en faveur de la marche et du vélo⁹ comme modes alternatifs à la voiture, en vantant leurs propriétés bénéfiques sur la santé. Ces modes, qui caractérisaient autrefois les déplacements des plus pauvres, deviennent les symboles d'une « conscience verte » et hygiénique, actuellement très en vogue. Une nouvelle fois, l'automobile redevient davantage qu'une simple modalité de déplacement : elle est érigée en objet d'assujettissement, qui aurait éloigné l'homme de sa condition « naturelle ».

⁸ Elles représentent 55% du total des émissions imputables au transport routier en 2004 (CITEPA, 2006).

⁹ En proposant de nouveaux services, comme « Vélib' » à Paris.

C'est pourquoi la dépendance à l'automobile est devenue ces dernières années une thématique importante dans la recherche en transports¹⁰. Toutefois, la notion de dépendance, polysémique, peut recouvrir des formes multiples. Par exemple, les résidents des zones rurales mal desservies par les transports en commun qui travaillent, s'approvisionnent, se soignent, etc., dans les zones urbaines proches sont à l'évidence plus dépendants de la voiture que les citadins sur place : il s'agit là d'une dépendance contextuelle. Mais aussi, lorsque la voiture devient le réflexe de la mobilité, un *habitus*, les agents qui font un usage routinier de l'automobile, quel que soit le déplacement, font potentiellement état d'un autre type de dépendance : celle à leurs habitudes. Autrement dit, ces agents peuvent avoir « pris le goût » d'utiliser la voiture au cours du temps (HIVERT, 2006).

Pour éviter la confusion des genres, cette dernière forme de dépendance est dans la théorie microéconomique nommée « addiction » (BECKER et MURPHY, 1988 ; BECKER et al., 1994). Si habitudes et addiction évoquent habituellement des comportements non réfléchis et inconscients, c'est au contraire dans un cadre de rationalité parfaite que BECKER et MURPHY (1988) développent un modèle qui considère l'évolution des goûts dans le temps, baptisé « modèle d'addiction rationnelle ».

Dans le second chapitre, nous effectuons une étude du comportement d'usage des automobiles à la disposition des ménages, en nous intéressant à leur kilométrage annuel parcourus, i.e. leur *automobilité*. Le cadrage théorique associé repose sur la microéconomie classique du consommateur, et les modèles de demande statique et dynamique sont abordés pour ainsi considérer la consommation individuelle de kilomètres. Nous visitons notamment le modèle d'addiction rationnelle, et dérivons sa forme fonctionnelle de demande. En expliquant la consommation courante simultanément par les consommations passée et future, nous voyons sous quelles hypothèses microéconomiques particulières les fonctions de demande standard et dynamique en ajustement courant partiel sont des cas particuliers du modèle d'addiction rationnelle. Pratiquement, sa mise en estimation oblige le suivi longitudinal individuel sur trois périodes d'observation au minimum : nous avons ainsi pu déployer le calque théorique du modèle d'addiction sur le panel

¹⁰ Particulièrement à la suite des travaux de NEWMAN et KENWORTHY (1989), qui ont rencontré un écho exceptionnel avec cette formule accrocheuse de « automobile dependence ».

pondéré Parc Auto 1999-2001. Les résultats font apparaître qu'on ne peut réfuter l'hypothèse microéconomique d'addiction concernant l'usage automobile. Quelques valeurs d'élasticité du kilométrage au prix des carburants dans un premier temps, puis au revenu des ménages ensuite, sont reportées, avec des résultats tout à fait plausibles. Ensuite, un éclairage est porté sur les différences géographiques, en confrontant les ménages des trois aires franciliennes et la Province, et en simulant l'effet d'un changement résidentiel sur leur mobilité. Enfin, nous voyons dans quelle mesure un changement des caractéristiques de motorisation (nombre d'automobiles, de permis, transitions essence-diesel) affecte l'automobilité des ménages français.

Les statistiques du second chapitre confirment l'évidence que l'usage automobile des ménages est en partie déterminé par le nombre de voitures à leur disposition. Leur niveau d'équipement prend donc à son tour une dimension écologique d'actualité. Son étude est complémentaire de celle de l'usage automobile dès lors que l'on souhaite comprendre et prévoir la demande de déplacement des ménages : elle doit ainsi intéresser la gestion de l'environnement, mais aussi celle des infrastructures routières, les constructeurs automobiles et les assurances.

Le second fondamental de la motorisation qui est exploré dans ce travail concerne le degré d'équipement automobile choisi par les ménages en France. L'analyse des niveaux de motorisation individuels est conduite dans le troisième chapitre, et exploite également les données de panel 1999-2001 de Parc Auto, déjà utilisées au cours du second chapitre. Une nouvelle fois, celles-ci se révèlent particulièrement adaptées pour aborder la dynamique comportementale. A l'échelle du ménage en effet, l'équipement automobile n'est pas fréquemment remis en question, et ne s'ajuste manifestement pas avec fluidité : les habitudes de mobilité (chapitre 2), la prise de décision puis la recherche d'informations sont quelques freins qui retiennent un ménage dans son état d'équipement, et lui demandent du temps avant de le modifier. De fait, le passé de motorisation des ménages pourrait être un facteur déterminant pour expliquer leur taux courant d'équipement, ce qui rend les données de panel indispensable à notre étude. Techniquement, notre analyse utilise un modèle probit ordonné dynamique avec hétérogénéité qui prend en considération trois états de motorisation : le non, le mono, et le multi-equipement. En spécifiant un modèle latent autorégressif simple, nous adaptons avantageusement quelques méthodes issues des modèles linéaires dynamiques pour exploiter et interpréter notre modèle catégoriel. Notamment, nous mesurons la

vitesse des ménages à ajuster leur niveau de motorisation. Egalement, nous évaluons pour un ensemble de ménages les effets d'un changement résidentiel vers une zone francilienne sur leurs probabilités de motorisation. Enfin, notre méthode permet d'évaluer, à court et long terme, les élasticités des probabilités d'équipement au revenu des ménages.

Les questions de l'usage et de la quantité du parc automobile à disposition des ménages ainsi traitées, le chapitre suivant fait place à la qualité du parc automobile. La pollution mesurée à la sortie du pot d'échappement est marquée par des différences qui tiennent aussi aux caractéristiques physiques des véhicules. A puissance, utilisation, et âge comparables, une voiture essence utilisant un supercarburant produit davantage de gaz à effet de serre¹¹ au kilomètre qu'une voiture diesel, mais rejette moins de particules dans l'atmosphère. Par ailleurs, la campagne actuelle de lutte contre le gaspillage énergétique (des ressources fossiles particulièrement), l'augmentation récente du prix des carburants et la sensibilisation médiatique aux conséquences d'accidents ont conduit les constructeurs à accélérer leurs efforts pour rendre les automobiles neuves moins énergivores au kilomètre, moins polluantes et plus sécurisées. Outre l'âge et le carburant utilisé, la puissance intervient : si les voitures les plus puissantes et les plus lourdes¹², qui caractérisent notamment le niveau haut de la gamme automobile, sont fréquemment les plus robustes et les plus sûres sur la route, elles sont aussi souvent plus énergivores que les véhicules classés en moyen et bas de gamme, et plus difficiles à garer.

Les automobilistes apparaissent donc très inégaux en fonction du type de voiture qu'ils acquièrent : inégaux devant le confort, la sécurité et dans la pollution qu'ils produisent... Favoriser l'entrée des véhicules propres dans le parc automobile, ou révéler les inégalités d'accès et de préférence des agents face aux différents segments de l'offre sont des sujets qui requièrent l'étude des choix d'acquisition.

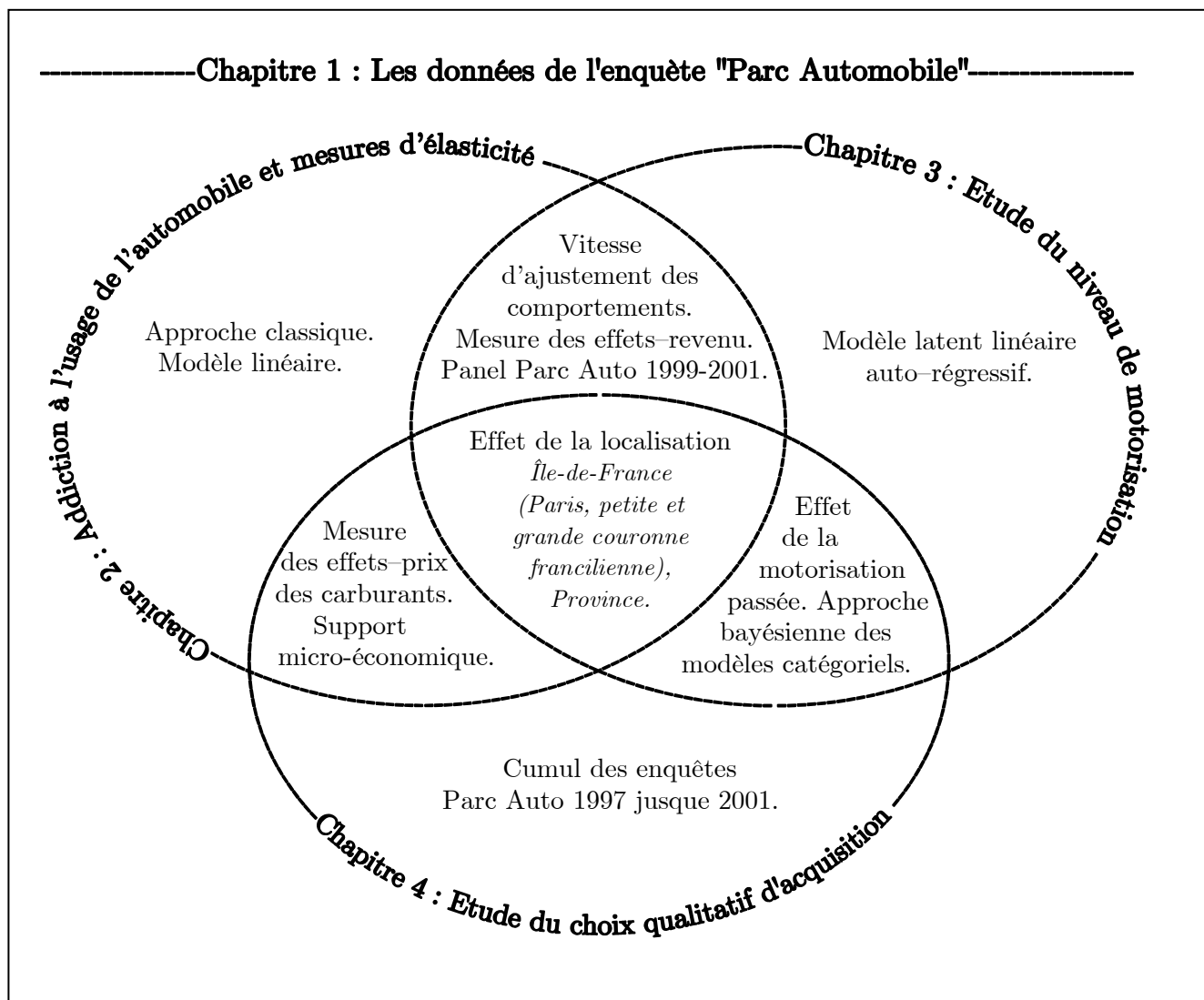
Nous conduisons cette étude au long du quatrième chapitre. En croisant les trois variables binaires du niveau de gamme (haut ou bas), du carburant (gazole ou supers), et de l'état à l'achat (neuf ou occasion), huit types de voitures sont considérées comme alternatives concurrentes de choix. Après avoir collecté dans les fichiers « voitures » de Parc Auto depuis 1997 jusqu'en 2001 toutes les observations

¹¹ Actuellement, le rapport technologique d'émission de CO₂ au kilomètre se situe aux environs de 3/4.

¹² Les « grosses cylindrées ».

d'automobiles acquises en cours d'année, puis identifié leur type, nous expliquons ces choix par les caractéristiques des conducteurs principaux, et l'indice des prix des carburants. Pour chaque voiture acquise, les vagues d'enquête Parc Auto contiennent des informations suffisantes permettant d'identifier, s'il y a lieu, le type de l'automobile précédente : nous les exploitons afin de tester la fidélité des agents aux caractéristiques de leur ancien véhicule, et conclure sur la dynamique temporelle des choix d'acquisition. Nous plaçons notre étude dans le cadre théorique de la microéconomie du consommateur tout d'abord, puis ajustons un modèle catégoriel de choix discrets non ordonnés pour caractériser le comportement d'acquisition individuel. Plus précisément, le modèle probit multinomial bayésien est estimé en utilisant une méthode de Monte Carlo par chaînes de Markov hybride (NOBILE, 1998). Nous exploitons les résultats obtenus en effectuant une analyse « *ceteris paribus* », afin de déterminer l'impact d'un changement de caractéristiques d'un agent moyen et l'effet de la hausse du prix des carburants sur la répartition des probabilités de choix (ou des parts des marchés) entre les différentes alternatives considérées. Nous notons ainsi quelques singularités d'équipement selon la zone de résidence des agents, les résidents franciliens vouant notamment une préférence plus forte pour les véhicules essence que les Provinciaux. Enfin, nous observons que la progression globale des prix des carburants marque un gain d'intérêt des acquéreurs pour les véhicules diesel, au détriment des voitures utilisant un supercarburant. En guise de résumé, le schéma de la figure 0.2 propose une vue d'ensemble de l'architecture de la thèse et reprend les traits d'analyse principaux et communs que nous aborderons au long des quatre chapitres suivants.

FIGURE 0.2 : Contenu et architecture de la thèse



Note de lecture : à l'intersection des bulles se trouvent les caractéristiques d'analyse communes aux chapitres concernés.

Chapitre 1 :

*Les données de l'enquête «Parc
Automobile»*

1.1 Contexte de l'enquête

Les données issues de l'enquête « Parc Automobile », dorénavant notée « Parc Auto », nous ont servi de support empirique tout au long de ce mémoire. Cette source procède d'une coopération verticale : elle est cofinancée par l'Agence De l'Environnement et de la Maîtrise de l'Energie (ADEME), le Comité des Constructeurs Français d'Automobiles (CCFA), la Fédération Française des Sociétés d'Assurances (FFSA) et la Direction de la Sécurité Routière (DSCR), elle est ensuite réalisée par l'institut d'études SOFRES auprès de ses panélistes « Métascope¹³», puis exploitée par l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) notamment.

L'ambition de l'enquête est la connaissance tant globale que précise du parc automobile français à la disposition des ménages résidents, en termes d'équipement et d'usage. Ainsi, Parc Auto constitue une source de données très ciblée qui fournit une description approfondie des attributs, de la qualité, et de l'utilisation des automobiles, mais aussi de leurs utilisateurs (individus et ménages).

Depuis 1976, cette enquête à vocation nationale est réalisée annuellement. Alors qu'elle était exclusivement destinée aux ménages dont les chefs de famille étaient français, elle s'est récemment ouverte aux familles étrangères (en 2002). Actuellement, elle est conduite en sondant par voie postale un échantillon aléatoire de 10000 ménages volontaires, dont le renouvellement observé est d'environ un tiers par an, et concerne prioritairement les « mauvais répondants ».

Si, globalement, deux questionnaires envoyés sur trois sont retournés¹⁴, ces réponses ne sont malheureusement pas aléatoires. En effet, le formulaire d'enquête s'avère rapide et facile à remplir pour les ménages petits (et non-motorisés), alors qu'il est contraignant et long pour les ménages de plus grande taille (et multi-motorisés). Les premiers sont donc sur-représentés (et les seconds, sous-représentés) dans le volume des réponses collectées.

Pour redresser cet échantillon des ménages répondants, une pondération individuelle est calculée par l'institut SOFRES après saisie. Concrètement, l'attribution aux ménages de poids s'effectue de telle façon que les distributions corrigées de variables socio-économiques classiques se calquent sur les distributions

¹³ <http://www.tns-sofres.com/sofres/tns-sofres-outils.php>.

¹⁴ Par exemple, 6872 en 1999, 6368 en 2000 et 6438 en 2001.

théoriques calculées par l'Institut National de la Statistique et des Etudes Economiques (INSEE) à partir des données nationales issues de l'« Enquête Emploi ». SOFRES retient cinq critères pour le redressement de l'échantillon répondant : la région de résidence, la taille d'agglomération de résidence, la PCS (Profession - Catégorie Sociale) du chef de famille, l'âge du chef de famille et le nombre de personnes dans le foyer.

Physiquement, une enquête annuelle se présente en deux fichiers électroniques séparés. Le premier fichier, dit « fichier-ménages », regroupe l'ensemble des caractéristiques des ménages répondants, et décrit notamment pour chacun sa tranche de revenu, sa localisation géographique (code de la commune de résidence, codages en zones d'urbanisation...), ainsi que des caractéristiques de ses membres (âge, sexe, possession de permis, catégorie professionnelle, etc...). Le second fichier, dit « fichier-voitures », fait un descriptif de leur motorisation en détaillant jusque trois automobiles (voitures particulières et petits utilitaires) à disposition des ménages répondants. Plus particulièrement, ce fichier renseigne pour chaque véhicule le kilométrage au compteur, le kilométrage annuel parcouru, la marque, le modèle, le type, l'âge, le niveau de gamme, l'état à l'acquisition, le carburant utilisé, la consommation moyenne au 100 km, les utilisateurs, etc... Un numéro d'identifiant unique et invariable dans le temps pour chaque ménage permet un appariement exact de ces deux types de fichier d'observations pour une année particulière, ainsi que le suivi longitudinal individuel pour les ménages répondant à plusieurs vagues d'enquête successives. Ce dernier cas préfigure le montage de panels.

1.2 Statistiques descriptives

Dans cette section, nous exposons brièvement quelques résultats descriptifs afin d'illustrer le contenu de l'enquête Parc Auto. Au long des prochains chapitres, nous aurons l'opportunité de présenter d'autres indicateurs descriptifs plus approfondis, et en rapport plus étroit avec les pôles étudiés.

A partir des vagues 1990, 1999, 2000 et 2001, et en utilisant les données démographiques nationales (population et nombre de ménages métropolitains) renseignées par l'INSEE, la table 1.1 permet de suivre l'évolution générale des comportements automobiles français au fil du temps.

TABLE 1.1 : Evolution des principaux indicateurs de mobilité et d'équipement automobile (France métropolitaine)

Année	1990	1999	2000	2001
Population totale (en millions d'hab.)	53.8	58.4	58.8	59.2
Nombre de ménages (en millions)	21.5	23.8	24.1	24.4
dont non-motorisés	5.1 (23.7%)	4.5 (18.8%)	4.7 (19.4%)	4.7 (19.3%)
dont mono-motorisés	10.9 (50.7%)	12.1 (50.7%)	12.0 (49.9%)	11.9 (48.9%)
dont multi-motorisés	5.5 (25.6%)	7.3 (30.5%)	7.4 (30.7%)	7.8 (31.8%)
Parc automobile (en millions d'unités)	22.9	27.9	28.2	28.9
dont parc essence	19.2 (84.0%)	17.3 (62.0%)	17.1 (60.7%)	16.9 (58.4%)
dont parc diesel	3.7 (16.0%)	9.8 (35.2%)	10.7 (37.9%)	11.6 (40.1%)
dont parc imprécis ou autres	0 (0%)	0.8 (2.8%)	0.4 (1.4%)	0.4 (1.5%)
Trafic automobile (en milliards de km)	300	386	382	403
dont parc essence	222	191	184	182
dont parc diesel	78	185	193	215
dont parc imprécis et autres	0	10	5	6
Kilométrage annuel (moyenne en km)				
par automobile	13208	13842	13573	13947
par automobile essence	11644	11068	10766	10811
par automobile diesel	20972	18774	18078	18555
par automobile imprécise ou autre	.	13156	13163	12790
par ménage	13953	16218	15865	16519
par personne	5576	6609	6503	6809
Nombre d'automobiles par ménage	1.065	1.172	1.172	1.185
Nombre d'automobiles par personne	0.425	0.477	0.479	0.488

Notes : les automobiles comprennent également les véhicules utilitaires légers.

Sources : INSEE, HIVERT (2001) pour la vague 1990 ; vagues 1999, 2000 et 2001 de l'enquête Parc Automobile.

Au cours de la décennie 1990, le volume du parc automobile en France a augmenté de 21.8%, correspondant à un taux de croissance annuel de +2.2%. Cette progression a très largement profité aux véhicules diesel qui ont vu la taille de leur parc être multipliée par 2.6 dans cette période, alors que le volume des automobiles essence a diminué de 10%. Le nombre de ménages progressant moins vite, cela s'est

traduit par une augmentation du nombre moyen de voitures par ménage, celui-ci passant de 1.06 en 1990 à 1.17 en 1999. Entre ces deux dates, la part et le nombre de ménages non-équipés a diminué, alors que celle et celui de ménages multi-motorisés a augmenté. Même si leur nombre a augmenté, la proportion de ménages mono-équipés s'est maintenue à 50.7%, et au début des années 2000, on compte une automobile environ pour deux français.

Cet essor de l'équipement dans les années 1990 s'est accompagné d'une hausse considérable du trafic automobile, de 10 milliards de kilomètres par an environ, pour atteindre 386 milliards de kilomètres en 1999. Le phénomène diesel a notablement contribué à augmenter le kilométrage annuel moyen des automobiles : parti de 13208 km en 1990, celui-ci a cru de 640 km en neuf ans.

Entre 1999 et 2001, le volume du parc automobile a continué de croître plus rapidement que les indicateurs démographiques, mais sur un rythme légèrement inférieur à celui observé au cours des années 1990 (+1.8% par an). Le nombre moyen d'automobiles à la disposition des ménages a ainsi progressé, de 1.17 à 1.19 véhicules. On retrouve également le processus déjà amorcé dans la décennie 1990 de la diésélisation du parc automobile : en volume, le diesel a cru de 18%, alors que le parc essence a continué de diminuer (-2.3%).

Malgré une diminution de l'utilisation moyenne des véhicules essence et diesel entre 1999 et 2001, c'est encore la progression du diesel dans le parc qui soutient le kilométrage annuel moyen par automobile. En effet, celui-ci progresse de 13842 à 13947 km, alors que les kilométrages moyens des voitures essence et diesel ont dans le même temps diminué de 257 et 219 km respectivement. Plus généralement, le trafic automobile en France est passé de 386 à 403 milliards de kilomètres entre 1999 et 2001.

En termes d'équipement, on remarque dans la table 1.1 que l'année 2000 est un point dont les indicateurs sont plus proches de ceux en 1999 qu'en 2001. Toutefois, l'année 2000 marque un fort retrait des indicateurs de mobilité par rapport à 1999. Le trafic total, les kilométrages annuels par ménage, par personne, et par voiture ont effectivement tous diminué, une raison tenant à l'augmentation exceptionnelle du prix des carburants entre 1999 et 2000 (+18%).

1.3 La mise en panel des données

A partir des vagues annuelles 1999 jusqu'à 2001 de l'enquête Parc Auto, nous avons notamment identifié un sous-échantillon statistique de 3010 ménages répondant successivement aux trois années d'enquête. Ce suivi longitudinal (c'est-à-dire temporel) d'une série d'observations transversales (les ménages) a permis l'organisation de ce sous-échantillon sous la forme d'un panel.

Néanmoins, la participation au panel des ménages répondant en 1999 n'est pas aléatoire. Notamment, on pourra vérifier avec l'exemple de questionnaire reporté en annexe que son remplissage par un ménage lui est d'autant plus fastidieux que le parc automobile à sa disposition est important. Les ménages multi-motorisés qui font l'effort de renvoyer un formulaire correctement renseigné en 1999 se souviennent de la difficulté de cet exercice en 2000 et certains s'en affranchiront, très logiquement, plus souvent qu'un ménage non-motorisé. Par ailleurs, l'organisme collecteur des données (SOFRES) « sanctionne » davantage les ménages multi-motorisés, qui peinent plus que les autres à renvoyer un formulaire pleinement renseigné, en ne les ré-interrogeant pas l'année suivante. De fait, la multi-motorisation doit être un facteur d'attrition sévère dans la construction de panels avec les vagues annuelles de Parc Auto, mais d'autres facteurs peuvent être avancés, comme la réticence des ménages à renseigner leur revenu annuel¹⁵.

1.3.1 Sélection et biais

Si la vague 1999 se voulait représentative des ménages français, le sous-échantillon des ménages panélistes¹⁶ ne l'était donc manifestement pas. Pour asseoir cette conviction, nous avons ajusté un modèle catégoriel de type probit à l'indicatrice de la participation au panel 1999-2001 des ménages de la vague initiale (1999). Quelques résultats sont reportés dans la table 1.2 qui suit. Ceux-là font notamment apparaître que, toutes choses égales par ailleurs, la probabilité qu'un ménage participe au panel diminue significativement avec l'augmentation du nombre d'automobiles à sa disposition, et globalement avec l'accroissement du revenu, comme attendu. De même, elle diminue très significativement lorsqu'il ne

¹⁵ En fait, les réticences d'un ménage le poussent aux non-réponses, alors que trop de non-réponses côté poussent SOFRES à l'évincer de l'enquête l'année suivante.

¹⁶ « Panélistes » au sens économétrique : ici les ménages à la fois présents dans les vagues 1999, 2000 et 2001 de l'enquête Parc Auto.

renseigne pas son revenu. Enfin, nous remarquons que les chances de sélection dans le panel diminuent avec le nombre de femmes dans le ménage (à nombre constant d'adultes), lorsque le chef de famille devient chômeur, et à mesure que son âge s'éloigne de 58 ans¹⁷.

TABLE 1.2 : Sélection des ménages de la vague 1999 de Parc Auto au panel 1999-2001

Variable	Coefficient	Variable	Coefficient
constante	-1.513 (-8.84)	Indicatrices de tranche de revenu annuel du ménage (réf. : $\leq 75\text{kF}$)	
# de voitures	-0.073 (-3.11)		
# d'adultes	-0.038 (-0.75)		
# de femmes	-0.134 (-3.84)		
âge du chef	0.065 (9.96)		
âge carré du chef/100	-0.056 (-8.90)	[75kF ; 100kF[-0.013 (-0.26)
Indicatrices du statut d'activité du CDF (ref. : occupé)		[100kF ; 125kF[-0.073 (-1.47)
		[125kF ; 150kF[-0.136 (-2.69)
		[150kF ; 175kF[-0.118 (-2.06)
retraité	0.076 (1.47)	[175kF ; 200kF[-0.166 (-2.95)
chômeur	-0.175 (-2.79)	[200kF ; 300kF[-0.165 (-3.13)
		$\geq 300\text{kF}$	-0.217 (-3.08)
		non réponse	-0.553 (-7.52)

Détails d'estimation : ajustement du modèle probit à la variable binaire valant 1 si le ménage en 1999 est retrouvé dans les fichiers Parc Auto 2000 et 2001, 0 sinon. Ratios de student entre parenthèses, 6872 observations. Notes : l'estimation comprend d'autres variables explicatives (nombre d'enfants par tranche d'âge, indicatrices géographiques) dont les coefficients souvent non significatifs ne sont pas reproduits ici.

Le caractère sélectif à l'entrée dans le panel étant avéré, le sous-échantillon des ménages en 1999 suivis jusqu'en 2001 n'est donc pas représentatif des ménages français de 1999. Des résultats statistiques issus de ces données brutes pourraient présenter un biais de sélection pouvant compromettre la justesse des analyses. C'est pourquoi, soucieux de préserver la représentativité initiale de la vague 1999 de Parc

¹⁷ Résultat de l'optimum de la fonction quadratique d'âge du chef de famille aux paramètres de la table 1.2.

Auto, nous avons utilisé une méthode de pondérations des observations afin de redresser le sous-échantillon des panélistes en 1999. Précisément, nous avons utilisé la méthode de pondération par les scores de propension, empruntée au document de BATTISTIN et al. (2000).

1.3.2 Redressement par la méthode des scores de propension

Supposons que nous disposions d'un échantillon A représentatif d'une population Ω dont on souhaite étudier un comportement Y . Admettons par ailleurs que nous ayons un échantillon B qui décrit ce comportement et qui possède un tronc commun de variables descriptives Z avec l'échantillon A , mais que nous savons ne pas être représentatif de la population Ω . La fonction de répartition de la variable d'intérêt Y dans B contrôlée par la distribution de Z dans A , que nous notons $F_{Y|B}^A$, est donné par :

$$F_{Y|B}^A = \int_Z F_{Y|Z,B} \times f_{Z|A} dZ = \int_Z F_{Y|Z,B} \times \left(\frac{f_{Z|A}}{f_{Z|B}} \right) \times f_{Z|B} dZ. \quad (1.1)$$

L'expression précédente définit la fonction de pondération :

$$w(Z) = \frac{f_{Z|A}}{f_{Z|B}}. \quad (1.2)$$

Pour une observation i particulière appartenant à l'échantillon B et ayant les caractéristiques $Z = Z_i$, nous calculons une pondération $w_i = w(Z_i)$. Son effet est d'appuyer (ou alléger) le poids des observations $i \in B$ qui ont des caractéristiques Z_i sous- (sur-) représentées par rapport à la source de contrôle A . En appliquant le théorème de Bayes, les poids individuels peuvent aussi être formulés ainsi :

$$w_i = w(Z_i) = \frac{f_{Z_i|A}}{f_{Z_i|B}} = \frac{(\Pr(A | Z_i) \times f_{Z_i}) / \Pr(A)}{(\Pr(B | Z_i) \times f_{Z_i}) / \Pr(B)} = \frac{1 - e(Z_i)}{e(Z_i)} \times \frac{\Pr(B)}{\Pr(A)}, \quad (1.3)$$

où $e(Z_i) = \Pr(B | Z_i)$ est le *score de propension*¹⁸ qui traduit la probabilité conditionnelle à ses caractéristiques Z_i qu'un ménage i appartienne à B . Puisqu'il n'est pas directement observable, il nous faut estimer et substituer $e(Z_i)$ par $\hat{e}(Z_i)$ à l'aide d'un modèle catégoriel simple (de type probit ou logit), où la variable dépendante vaut 1 si $i \in B$, 0 si $i \in A$. L'estimation de ces scores de propension permet donc de calculer le nouveau système des pondérations w_i qui rend l'échantillon B statistiquement représentatif¹⁹ de l'échantillon A .

Pratiquement dans notre cas d'application, nous avons donc pris pour A la vague 1999 de Parc Auto, représentative des ménages français en 1999, et pour B l'échantillon identifié des ménages panélistes en 1999. Nous avons ensuite calculé un poids w_i de redressement pour chaque ménage panéliste en utilisant une modélisation probit pour estimer les scores de propension. Avec ces pondérations, nous disposons ainsi d'un échantillon redressé de ménages, représentatifs des ménages français en 1999, et dont nous avons le suivi longitudinal jusqu'en 2001. L'imputation de quelques valeurs par interpolation ou extrapolation des valeurs passées et/ou futures a clôt la formation du panel long de trois ans²⁰, composé de 3010 ménages individuellement pondérés, soit 9030 points.

¹⁸ L'appellation d'origine nous vient de RUBIN et ROSENBAUM (1983).

¹⁹ Pour vérifier le bon fonctionnement de cette méthode et montrer qu'il n'y a plus de sélection des ménages selon leurs caractéristiques Z avec les pondérations w_i , il suffit de réestimer le modèle catégoriel (probit ou logit) en considérant ces poids, et vérifier que tous les coefficients (exceptée la constante qui traduit des tailles d'échantillon différentes entre A et B) sont cette fois non significatifs.

²⁰ La variable la plus touchée par ces non réponses est le revenu annuel, mais ce phénomène reste très marginal pour les panélistes. Les quelques ménages n'ayant jamais renseigné leur revenu annuel entre 1999 et 2001 ont été supprimés.

Chapitre 2 :

*Addiction à l'usage de l'automobile et
mesures d'élasticités*

2.1 Introduction

Les préoccupations environnementales croissantes liées au réchauffement climatique mettent l'homme au banc des accusés. Le développement de ses activités s'est accompagné d'un rejet croissant dans l'atmosphère des gaz à effet de serre qui sont, sous l'action de la réflexion infrarouge des rayons solaires sur la surface du globe terrestre, en partie responsables de la hausse du thermomètre. Les conséquences écologiques à terme prévoient un bouleversement irréversible de l'écosystème, dans lequel la place de l'homme est incertaine et controversée. Soucieuse de sa préservation, la ratification du protocole de Kyoto (1997) par quelques pays dont la France marque une prise de conscience et une volonté communautaire de réduire les émissions de ces gaz, avec des objectifs ambitieux, pourtant jugés nécessaires.

Plus particulièrement, c'est l'activité de transport qui est visée comme étant une principale source génératrice de ces gaz, faisant notamment de l'automobile « l'instrument du délit ». La voiture est aussi impliquée dans bien d'autres affaires, plus locales cette fois, telles que la pollution atmosphérique, la sécurité routière, les nuisances sonores, la dégradation du paysage... mais la voiture permet aux agents une plus large maîtrise de l'espace et du temps, et donc un meilleur accès à l'emploi, aux services, aux loisirs, faisant de l'automobile un thème de débat très controversé.

Au cœur de toutes ces préoccupations particulières, la question de l'usage automobile et de son intensité émerge très nettement. L'« *automobilité*²¹ », qui désigne les comportements périodiques de mobilité en automobile des agents, s'impose donc comme un sujet majeur de réflexion.

La conduite de politiques pertinentes pour orienter ces comportements nécessite de connaître l'impact des instruments agissant sur ces déplacements parcourus en voiture individuelle et donc la manipulation d'un modèle descriptif du comportement d'usage. Sa connaissance est autrement utile pour les pouvoirs publics : en permettant la prévision, un modèle de trafic automobile peut servir la décision d'investissement dans les infrastructures routières, en complémentarité avec les modèles décrivant le niveau d'équipement en automobiles. Dans ce chapitre, nous proposons une structure en deux étapes pour caractériser l'usage de l'automobile en France : après avoir observé le comportement binaire de sélection

²¹ Mot-valise né du télescopage entre « automobile » et « mobilité ».

des ménages entre motorisation et non-motorisation, l'automobilité annuelle des ménages équipés est modélisée selon différentes spécifications, déterminées sur la base du support microéconomique. Ainsi, un modèle statique de demande de kilomètres en automobile est d'abord considéré, notamment pour choisir la variable de prix théoriquement pertinente. Dans un contexte de fortes variations du prix des carburants usuels transformés du pétrole - essence et gazole - qui alimentent encore la quasi intégralité du parc automobile français en 2000, et dans une perspective incertaine de leur évolution (liée à une demande croissante de produits pétroliers²², à l'hypothèse d'un tarissement à terme des ressources fossiles, et à une évolution incertaine des capacités d'offre et d'extraction de l'or noir notamment), un objectif du chapitre concerne la mesure des effets-prix sur le kilométrage des ménages. Plus particulièrement, un éclairage sur l'hétérogénéité de leurs sensibilités selon la zone résidentielle, notamment en Île-de-France, est tenté. L'effet d'autres variables explicatives, comme les variations de revenu est également considéré. Enfin, la disposition en panel des observations sur trois années (1999-2001), qui assure le suivi longitudinal des comportements individuels, permet l'utilisation de spécifications dynamiques, avantageuses pour distinguer les effets usuels de court et long terme.

Dans ce chapitre, les modèles dits « d'addiction myope » et « d'addiction rationnelle » sont étudiés puis ajustés afin de révéler et caractériser le type de dépendance des ménages à l'utilisation de leur parc automobile. En appliquant ces modèles, généralement utilisés pour décrire la consommation de cigarettes, de drogues ou d'alcool, on pourra vérifier la validité – au moins économique – des propos de DUPUY (1999) :

“... automobile dependence means that as individuals, we cannot live without cars, just as a smoker cannot live without cigarettes, and a drug addict without drugs” (DUPUY, 1999)

Dans la citation suivante, WICKHAM et al. (2002) osent également le rapprochement de l'usage automobile à celui d'une consommation addictive de drogue, dont on peut faire une relecture microéconomiste (entre parenthèses), motivant encore davantage l'emploi et le test de modèles d'addiction :

²² Par la Chine notamment.

“Car (use) dependence can be understood through the metaphor of drug dependency (addiction). Heroin or even nicotine addiction is in part a matter of (rational) choice. I choose to shoot up, I choose to smoke a cigarette. But as I continue to do this, my body (utility function) changes, it becomes restructured, it needs the drug (addictive good), it cannot do without it. Furthermore, the ‘need’ (addictive good marginal utility) escalates – the body (optimal bundle of goods) requires more and more of the drug (addictive goods).... the same applies to car (use) dependency” (WICKHAM et al., 2002).

Le chapitre est organisé de la façon suivante. Dans un premier temps, les modèles théoriques de demande (classique et d'addiction) sont examinés, et les spécifications économétriques sont dérivées. Les mesures d'élasticités sont ensuite explicitées dans une seconde partie. Après avoir observé quelques statistiques descriptives du panel Parc Automobile 1999-2001 (troisième partie), nous présentons dans un quatrième volet les difficultés et les résultats d'estimation des modèles de demande, les mesures d'élasticités associées et celles d'effets marginaux, avec un éclairage plus prononcé sur les comportements franciliens. Enfin, nous procédons dans la dernière partie à une revue de littérature afin d'apprécier la proximité de nos résultats avec ceux de travaux empiriques voisins déjà conduits.

2.2 Supports microéconomiques de la demande de kilomètres

2.2.1 Le modèle statique de demande de kilomètres

Dans la théorie néoclassique de la demande, les modèles de consommation statiques supposent la séparabilité temporelle de la fonction d'utilité des agents : à la période courante, ceux-ci choisissent leur panier optimal de biens qui maximise leur satisfaction dans un contexte où leur sont donnés comme contraintes le système de prix et leur dotation monétaire pour la dépense. L'agent ne peut constituer de stock et renouvelle sa demande de biens au début de chaque période de consommation. Sa fonction de demande (dite « marshallienne ») pour un bien particulier peut être obtenue depuis sa fonction d'utilité indirecte grâce à l'identité

de Roy (VARIAN, 1995, p.110), la quantité de bien demandée par l'agent étant alors une fonction du système de prix et de sa dotation allouée pour la dépense.

Dans les lignes suivantes, nous proposons un modèle statique de demande de transport automobile : il repose sur la possibilité pour un agent d'arbitrer entre une consommation d'un bien composite et une consommation de kilomètres.

A partir de son lieu résidentiel, on suppose qu'un agent rationnel peut effectuer K déplacements automobiles, caractérisés par des destinations et des distances différentes. On retient les hypothèses classiques habituelles en microéconomie d'une croissance de l'utilité et d'une décroissance de l'utilité marginale avec les quantités, ici le nombre de déplacements effectués vers ces destinations.

Autrement dit, on considère que, toutes choses égales par ailleurs, la courbe mesurant la satisfaction conditionnelle retirée par l'agent de s'être rendu en la destination k , notée U_k , présente une allure croissante et concave avec les quantités q_k :

$$\frac{\partial U_k}{\partial q_k} > 0 ; \left(\frac{\partial^2 U_k}{\partial q_k^2} = \frac{\partial U_k'}{\partial q_k} \right) < 0 . \quad (2.1)$$

Chaque déplacement en automobile vers la destination k peut être effectué par l'agent en contrepartie d'une dépense ayant un coût unitaire p_k , supposé constant. On considère par ailleurs un bien composite, noté X , synthétique de l'ensemble des autres biens consommables dans l'économie et dont le prix p_X est usuellement fixé au numéraire. Si la fonction d'utilité conditionnelle du bien X présente les mêmes caractéristiques de croissance concave avec les quantités consommées, alors les courbes d'indifférence de l'agent sont convexes et la courbe enveloppe de son utilité est concave. L'agent rationnel qui a maximisé son bien-être sous sa contrainte budgétaire a choisi son panier optimal de biens $P^* = \{q_1^*, \dots, q_K^*, q_X^*\}$. S'il n'y a aucune solution en coin, son programme d'optimisation implique :

$$\frac{U_k'(P^*)}{p_k} = \frac{U_X'(P^*)}{p_X} = U_R'(P^*), \quad \forall k = (1, \dots, K), \quad (2.2)$$

avec $U_R'(P^*)$, la valeur de l'utilité marginale de la richesse au panier de biens optimal. Si au contraire son panier optimal comporte des non déplacements vers

une destination particulière k , ou des non consommations du bien composite (c'est-à-dire pour $q_k^* = 0$ ou $q_X^* = 0$), alors le choix de l'agent respecte l'inégalité suivante :

$$\frac{U'_k(P^*)}{p_k} \leq U'_R(P^*), \text{ ou } \frac{U'_X(P^*)}{p_X} \leq U'_R(P^*). \quad (2.3)$$

Les (in)égalités (2.2) et (2.3) résultent de la rationalité de l'agent qui dépense numéraire par numéraire jusqu'à épuisement de son allocation R dans l'achat du bien composite et des déplacements, qui lui procurent successivement et de façon décroissante les niveaux de satisfaction marginales les plus élevés.

Dans notre étude, nous posons que le coût pour l'agent de chaque déplacement aller-retour depuis son domicile vers une destination k est une simple fonction multiplicative de la longueur totale de trajet en kilomètres, notée $2km_k$ ²³, et du prix du kilomètre p_{km} . Ce prix est le coût marginal de la production d'un kilomètre supplémentaire avec son automobile : dans une version simplifiée il correspond au produit de la consommation moyenne en carburant de son véhicule, cmc , et du prix du carburant utilisé pour l'alimenter, p_{carb} . Formellement :

$$p_{km} = p_{carb} \times cmc. \quad (2.4)$$

Le coût supporté par l'agent pour la production du transport automobile l'emmenant à la destination k , puis le ramenant à son domicile devient ainsi :

$$p_k = 2km_k \times p_{km}. \quad (2.5)$$

En divisant les numérateurs et dénominateurs des utilités marginales relatives des déplacements k en (2.2) par leur distance totale $2km_k$, celles-ci deviennent :

$$\frac{U'_k(P^*)/2km_k}{p_k/2km_k} = \frac{U'_{km_k}(P^*)}{p_{km}} = \frac{U'_{km}(P^*)}{p_{km}} ; \forall k \in [1; K]. \quad (2.6)$$

Cette double égalité mérite un commentaire détaillé. Dans la première partie de (2.6), la division au numérateur a pour effet de répartir l'utilité marginale des

²³ Le facteur 2 tient au fait que l'on considère l'aller et le retour.

déplacements sur chaque kilomètre qu'ils comptent. Sous l'hypothèse classique de divisibilité des biens²⁴, et puisque les déplacements sont caractérisables par une distance kilométrique, il devient équivalent d'exprimer le panier optimal en quantités de déplacements ou en kilomètres totaux parcourus vers chacune des K destinations. Dès lors, la division exprime l'utilité marginale d'un kilomètre parcouru vers la destination k . Au dénominateur, la division permet de rendre le prix indépendant de la destination : pour un prix kilométrique constant, la seconde partie de l'égalité traduit que l'optimalité du choix de l'agent n'est assurée que lorsque les utilités marginales d'un kilomètre vers les K destinations sont toutes égales.

Finalement, cette manipulation réduit l'ensemble des K déplacements à K « paquets » de kilomètres dans lesquels le consommateur va puiser successivement afin de maximiser son utilité, en payant un prix unique pour chacun. Autrement dit, l'agent va produire et consommer des kilomètres qu'il répartira entre les différentes destinations, et qui lui coûtent un prix constant : p_{km} . Ainsi, le kilomètre peut être considéré comme un bien de consommation classique. A condition d'utiliser son prix (et non celui du carburant p_{carb} directement), l'estimation empirique de sa fonction de demande constitue un travail théoriquement fondé, au sens microéconomique.

Cette fonction de prix pourrait sembler *a priori* légère à certains égards, puisqu'elle ne tient pas compte des coûts fixes liés à l'amortissement de la voiture ou encore aux coûts perdus d'assurance du véhicule. Mais ces coûts n'interviennent théoriquement qu'en amont, au moment de la décision de participer ou non au « cercle » des usagers de l'automobile. Par ailleurs, la majorité des autres coûts sont également fixes pour les agents, comme certains coûts engendrés par l'entretien de leur automobile (vidange et changements des filtres annuels, contrôle technique biannuel obligatoire, remboursements mensuels fixes pour une auto achetée à crédit par exemple). Puisque les automobilistes sont à la fois producteurs et consommateurs de kilomètres, on peut envisager leur optimum avec la théorie microéconomique du producteur. Considérons une fonction de coût total Ct incorporant ces coûts fixes, notés Cf :

$$Ct(km) = Cf + Cvt(km), \quad (2.7)$$

²⁴ Comme pour une majorité de biens de consommation, la validité empirique de cette hypothèse théorique est très limitée.

avec Cvt la fonction linéaire du coût variable total, définie par:

$$Cvt(km) = km \times p_{km} . \quad (2.8)$$

L'optimum du producteur km^* est atteint lorsqu'il y a l'égalité entre la recette marginale et le coût marginal de production :

$$Rm(km^*) = Cm(km^*) . \quad (2.9)$$

Le coût marginal est obtenu en dérivant la fonction de coût total, et correspond ici au prix unitaire du kilomètre :

$$Cm(km) = p_{km} . \quad (2.10)$$

Il s'agit bien d'une constante indépendante des coûts fixes Cf . La recette marginale (ou la disposition marginale à payer) correspond à l'équivalent monétaire de l'utilité marginale du kilomètre :

$$Rm(km) = \frac{U'_{km}(P)}{U'_R(P)} \quad (2.11)$$

Les relations (2.9), (2.10), (2.11) permettent notamment de retrouver les égalités (2.2) et (2.6). $Rm(km)$ est une fonction décroissante avec les kilomètres compte tenu de l'hypothèse de décroissance de l'utilité marginale avec les quantités consommées. On détermine « l'output » optimal km^* à l'aide du graphique bien connu de la solution du producteur (Figure 2.1).

La surface comprise entre les droites Rm et Cm jusqu'en km^* (figure 2.1, triangle « abc ») mesure le surplus, ou profit, brut de l'agent. Le surplus net correspond au surplus brut amputé du coût fixe, graphiquement représenté par la surface du rectangle « bcde » sur la figure 2.1. Nous constatons bien que la valeur d'output optimal reste inchangée quel que soit le niveau des coûts fixes, tant qu'il existe un surplus net positif. Si celui-ci est négatif, l'agent ne produira pas de kilomètres et choisira d'être non-motorisé tout comme un producteur choisira de ne pas entrer sur un marché si son profit est au final négatif. Ainsi, le coût pertinent à utiliser pour modéliser le kilométrage des agents motorisés est bien celui du coût

marginal, ici le coût constant engendré par la production d'un kilomètre²⁵. Ce modèle statique et les modèles dynamiques qui suivent peuvent dès lors être mis en œuvre afin d'observer la réactivité moyenne des agents aux variations des déterminants de leur consommation individuelle de kilomètres.

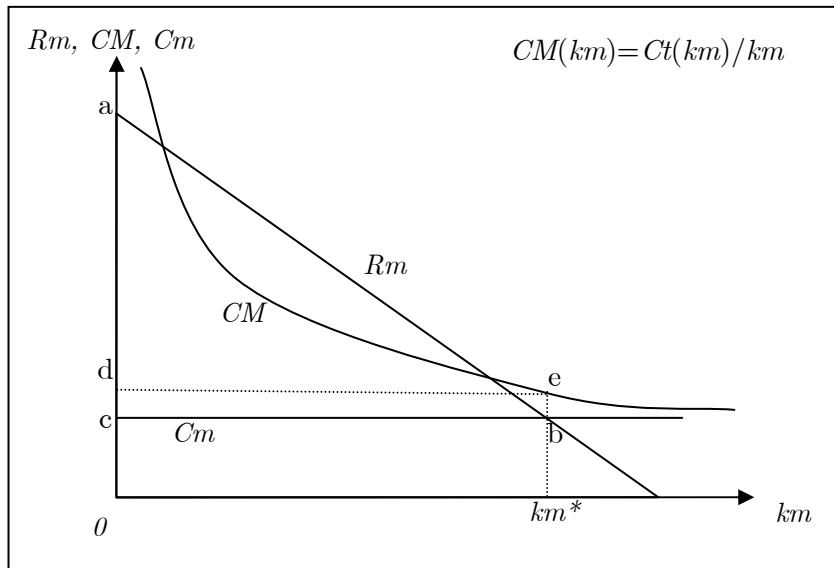


FIGURE 2.1 : Détermination de l'output optimal de kilomètres dans la théorie du producteur

2.2.2 Modèles dynamiques de demande : les modèles d'addiction myope et rationnelle de Becker

2.2.2.1 Description

Depuis BECKER et MURPHY (1988), l'addiction d'un consommateur pour un bien est révélée si, toutes choses égales par ailleurs, l'augmentation de sa consommation passée vient augmenter significativement sa consommation courante. Le modèle d'addiction considéré par BECKER, GROSSMAN et MURPHY (1994) établit que le niveau d'utilité U_t retirée par un individu au cours de la période t dépend

²⁵ Il peut exister d'autres coûts directement reliés au niveau de kilométrages tels que ceux de réparations. Leur nature aléatoire difficilement traitable et les objectifs du chapitre font qu'ils ne seront pas considérés.

de la quantité consommée de deux types de biens : une quantité X_t d'un bien composite X , une quantité C_t d'un bien addictif C , ainsi que de l'impact d'une série de variables relatives au cycle de vie, potentiellement non mesurées, et notées e_t . On différencie les deux types de consommation en posant que l'utilité courante U_t dépend également de l'importance des quantités de bien C consommées dans le passé. Selon les auteurs, cette information est synthétisée dans la formulation d'un stock de capital addictif, noté S_t , alimenté à chaque période par le niveau de consommation antérieure C_{t-1} . Dès lors, l'utilité courante d'un agent doit également dépendre de ce stock. Nous retenons l'expression de BECKER et al. (1994) qui est la plus commune dans les travaux de modélisation traitant du phénomène d'addiction : $S_t = C_{t-1}$. Cependant, d'autres types de formulation pour la constitution du stock sont envisagés²⁶ (BECKER et MURPHY, 1988).

Le problème économique de l'agent devient donc le suivant : il doit maximiser son utilité intertemporelle, ou la somme actualisée par un taux de préférence pour le présent (ou un taux de dépréciation) ρ de ses niveaux d'utilité dans le temps. En considérant que l'agent est rationnel, qu'il a une durée de vie illimitée, et que sa consommation de bien addictif n'influe pas sur ses revenus, alors son programme de maximisation d'utilité s'écrit :

$$\text{Max} \sum_{t=1}^{\infty} B^{t-1} U_t(C_t, C_{t-1}, X_t, e_t), \quad (2.12)$$

avec $B = (1 + \rho)^{-1}$. Le bien composite X utilisé par Becker est le numéraire. L'auteur fait aussi l'hypothèse simplificatrice que le taux de préférence pour le présent équivaut au taux d'intérêt de l'économie. Enfin, l'agent est soumis au respect de son équilibre budgétaire intertemporel et à une condition initiale de départ pour C , en $t = 0$:

$$C_0 = C^0 ; A_0 = \sum_{t=1}^{\infty} B^{t-1} (X_t + P_t C_t), \quad (2.13)$$

avec A_0 , la valeur actualisée de la richesse de l'agent, et P_t le prix du bien addictif en t . Sous l'hypothèse que la fonction d'utilité de l'agent est de forme quadratique

²⁶ Notamment, sur la base d'un taux de dépréciation τ du stock passé et réalimenté par la consommation passée du bien addictif : $S_t = (1 - \tau) S_{t-1} + C_{t-1}$. La formulation la plus courante $S_t = C_{t-1}$ correspond donc au cas particulier d'une dépréciation totale : $\tau = 100\%$.

en l'ensemble de ses arguments C_t, C_{t-1}, X_t, e_t , puis en résolvant les conditions de premier ordre qui maximisent l'utilité intertemporelle de l'agent, BECKER et al. (1994) aboutissent à la fonction de demande pour le bien addictif. Formellement :

$$U_t(C_t, C_{t-1}, X_t, e_t) = \alpha_C C_t + \alpha_S C_{t-1} + \alpha_X X_t + \alpha_e e_t + \frac{\alpha_{CC}}{2} C_t^2 + \frac{\alpha_{SS}}{2} C_{t-1}^2 + \frac{\alpha_{XX}}{2} X_t^2 + \frac{\alpha_{ee}}{2} e_t^2 + \alpha_{CS} C_t C_{t-1} + \alpha_{CX} C_t X_t + \alpha_{Ce} C_t e_t + \alpha_{SX} C_{t-1} X_t + \alpha_{Se} C_{t-1} e_t + \alpha_{Xe} X_t e_t. \quad (2.14)$$

La solution du programme de maximisation de l'agent sous sa contrainte budgétaire suit l'écriture du lagrangien L :

$$L = \sum_{t=1}^{\infty} B^{t-1} (U_t(C_t, C_{t-1}, X_t, e_t)) + \lambda \left(A_0 - \sum_{t=1}^{\infty} B^{t-1} (X_t + C_t P_t) \right). \quad (2.15)$$

On résout ce problème de maximisation sous contrainte en annulant les dérivées partielles de L :

$$\frac{dL}{dC_t} = B^{t-1} \frac{dU_t(C_t, C_{t-1}, X_t, e_t)}{dC_t} + B^t \frac{dU_{t+1}(C_{t+1}, C_t, X_{t+1}, e_{t+1})}{dC_t} - \lambda B^{t-1} P_t = 0, \quad (2.16)$$

$$\frac{dL}{dX_t} = B^{t-1} \frac{dU_t(C_t, C_{t-1}, X_t, e_t)}{dX_t} - B^{t-1} \lambda = 0, \quad (2.17)$$

avec λ le multiplicateur de Lagrange, qui correspond à l'utilité marginale de la richesse intertemporelle A_0 . En simplifiant par B^{t-1} , il découle de (2.17) et de (2.14) que :

$$\lambda = \frac{dU_t(C_t, C_{t-1}, X_t, e_t)}{dX_t} = \alpha_X + \alpha_{XX} X_t + \alpha_{CX} C_t + \alpha_{SX} C_{t-1} + \alpha_{Xe} e_t, \quad (2.18)$$

puis en exprimant X_t :

$$X_t = \frac{\lambda - (\alpha_X + \alpha_{CX} C_t + \alpha_{SX} C_{t-1} + \alpha_{Xe} e_t)}{\alpha_{XX}}. \quad (2.19)$$

Après la simplification par B^{t-1} , il vient analytiquement de (2.16) l'égalité suivante :

$$\begin{aligned} \lambda P_t = & \alpha_C + \alpha_{CC}C_t + \alpha_{CS}C_{t-1} + \alpha_{CX}X_t + \alpha_{Ce}e_t \\ & + B(\alpha_S + \alpha_{CS}C_{t+1} + \alpha_{SS}C_t + \alpha_{SX}X_{t+1} + \alpha_{Se}e_{t+1}). \end{aligned} \quad (2.20)$$

En remplaçant X_t dans (2.20) par son expression donnée en (2.19), puis en exprimant C_t , on aboutit à la fonction de consommation, ici exprimée sans constante, de BECKER et al. (1994) :

$$C_t = \theta C_{t-1} + \theta B C_{t+1} + \theta_1 P_t + \theta_2 e_t + \theta_3 e_{t+1}, \quad (2.21)$$

$$\text{avec :} \quad \theta = -D^{-1}(\alpha_{XX}\alpha_{CS} - \alpha_{CX}\alpha_{SX}), \quad (2.22)$$

$$\theta_1 = D^{-1}\lambda\alpha_{XX}, \quad (2.23)$$

$$\theta_2 = -D^{-1}(\alpha_{XX}\alpha_{Ce} - \alpha_{CX}\alpha_{Xe}),$$

$$\theta_3 = -D^{-1}(\alpha_{XX}\alpha_{Se} - \alpha_{SX}\alpha_{Se}),$$

$$\text{et sachant :} \quad D = (\alpha_{CC}\alpha_{XX} - \alpha_{CX}^2) + B(\alpha_{SS}\alpha_{XX} - \alpha_{SX}^2). \quad (2.24)$$

La quantité courante demandée pour la consommation du bien addictif C formulée en (2.21) est au final une fonction des demandes passée et future (C_{t-1} et C_{t+1}), du prix courant P_t , et des variables liées au cycle de vie de l'agent e_t et e_{t+1} .

Dans, l'équation (2.24), D représente la somme actualisée des mineurs d'ordre deux du Hessian de la fonction d'utilité en (2.14) pour les quantités consommées de biens C et X . Par hypothèse microéconomique usuelle, la fonction d'utilité U est concave. Il en résulte que D est nécessairement positif : $D > 0$. La concavité de U_t implique également que les mineurs de son Hessian au premier ordre soient négatifs ; et donc que $\alpha_{XX} < 0$. De plus, l'utilité marginale de la richesse intertemporelle λ étant positive, il vient que le coefficient θ_1 exprimé en (2.23) est négatif : $\theta_1 < 0$. Les niveaux de consommation passé et courant du bien C sont dits complémentaires lorsque α_{CS} est strictement positif. Dans ce cas de figure, l'utilité marginale courante de l'agent, retirée par la consommation d'une quantité additionnelle du bien C en t et notée U'_{C_t} en (2.25), est une fonction croissante de C_{t-1} :

$$U'_{C_t} = \frac{dU_t}{dC_t} = \alpha_C + \alpha_{CC}C_t + \alpha_{CS}C_{t-1} + \alpha_{CX}X_t + \alpha_{Ce}e_t. \quad (2.25)$$

Ainsi, la quantité C_{t-1} et le coefficient α_{CS} « relèvent » d'autant plus la satisfaction de l'agent retirée par la consommation marginale du bien C qu'ils sont positifs. Par analogie avec le concept du « *learning by doing* », l'agent aura d'autant plus *appris* à apprécier la consommation du bien C (U'_{C_t}) qu'il aura *pratiqué* cette consommation dans le passé (C_{t-1}), et que la vitesse d'apprentissage (α_{CS}) est élevée. La complémentarité temporelle des consommations du bien C est la marque de l'addiction. Si l'on considère de plus que le stock de capital addictif n'influence pas l'appréciation courante du bien composite (i.e pour $\alpha_{SX} = 0$), alors l'addiction entraîne que $\theta > 0$ en (2.22). Ainsi, l'ajustement des données sur le modèle de consommation (2.21) révèle un comportement addictif si l'importance des consommations passées induit un « renforcement » de la consommation courante. Autrement dit, et à toutes autres choses égales, la significativité empirique du coefficient θ traduit l'addiction à la consommation du bien C . Plus θ est positif, plus le renfort est important, et plus le degré d'addiction est élevé. Son estimation en (2.21) permet ensuite d'obtenir par déduction une estimation de B , puis une estimation du taux de substitution intertemporel ρ .

Les effets sur la consommation courante de chocs sur les consommations passée et future sont déduits à l'aide des racines caractéristiques solutions de l'équation homogène de la fonction de consommation du modèle d'addiction rationnelle. Celle-ci s'écrit :

$$\theta X^2 - X + \theta B = 0. \quad (2.26)$$

Les racines caractéristiques de l'équation (2.26) sont :

$$\varphi_1 = \frac{1 - \sqrt{1 - 4\theta^2 B}}{2\theta}, \quad (2.27)$$

$$\varphi_2 = \frac{1 + \sqrt{1 - 4\theta^2 B}}{2\theta}. \quad (2.28)$$

En (2.27) et (2.28), φ_1 mesure l'effet sur la consommation courante d'un choc sur la consommation future, et φ_2^{-1} mesure l'effet sur la consommation courante d'un choc sur la consommation passée. Toutes les élasticités du modèle

d'addiction rationnelle peuvent être exprimées en fonction de ces racines caractéristiques²⁷.

L'apport de BECKER et al. (1994) est plutôt conséquent. Leur fonction de demande dynamique résume les fonctions usuelles, statique et dynamique auto-régressive, à des cas très particuliers de la formulation (2.21). En effet, une équation de demande statique est opérante lorsque le degré d'addiction effectif est nul ($\theta = 0$). De même, une fonction de demande dynamique ne faisant apparaître qu'une variable endogène retardée comme facteur explicatif n'est pertinente dans le cadre de l'addiction rationnelle que lorsque les agents consomment au « jour-le-jour », mais avec un effet mémoire : ceux-là ignorent les effets futurs de leur consommation actuelle et rendent compte d'un taux de préférence intertemporel ρ infiniment élevé (lorsque $B = 1/(1 + \rho) = 0$). En réagissant de façon mécanique aux consommations passées, sans considérer les effets futurs de leur consommation actuelle, les agents témoignent d'une forme particulière d'addiction. Pour la caractériser, BECKER et al. (1994) utilisent le qualificatif « myope ».

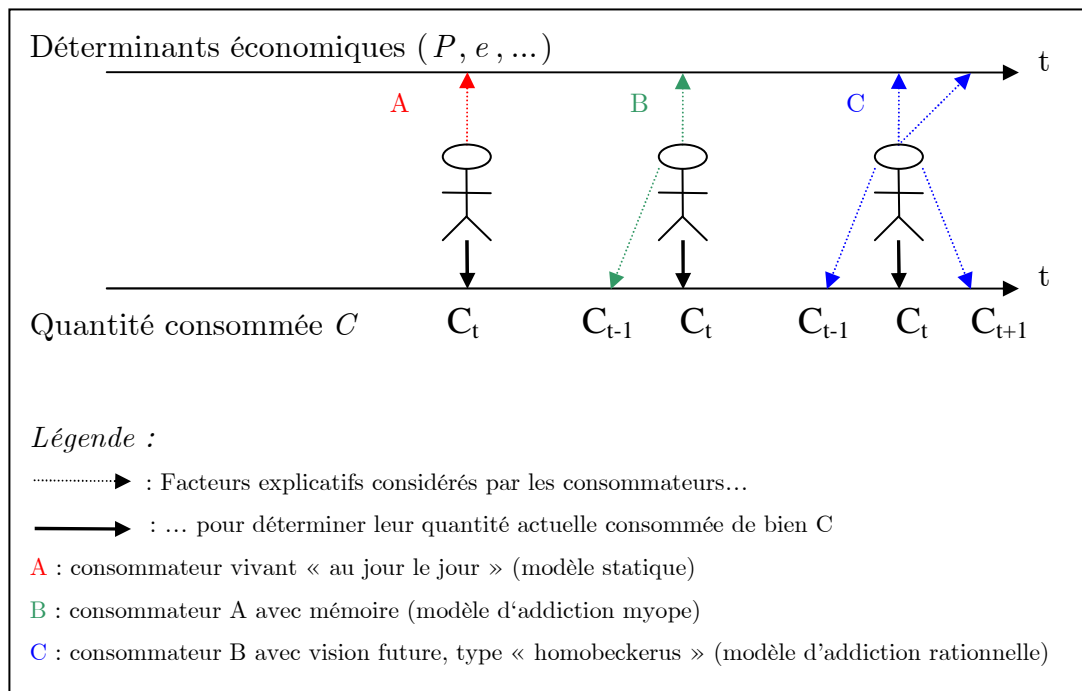


FIGURE 2.2 : Trois types de consommateurs, trois modèles de consommation

²⁷ Nous les formulerons plus tard, dans la section 2.3.4.

La figure 2.2 schématise les différents types de consommateurs visités jusqu'ici. Parmi ceux-là figure notamment celui que nous avons baptisé « l'homobeckerus ». C'est en effet une constante immuable de départ chez Becker de considérer en postulat microéconomique de ses recherches appliquées que l'agent est du type maximisateur d'utilité intertemporelle sous contrainte de ses revenus actualisés et de ses stocks de capital, définissant ce que LAWSON (2004) appelle le « style économique de l'école de Chicago ». A cet égard, et dans un document au titre accrocheur « *Gary Becker and the Quest for the Theory of Everything* », l'auteur estime que Becker a une approche monothéorique préformatée des phénomènes humains :

« (he extends) an existing theoretical model (individual utility maximisation constrained by budget level and current capital stocks) to cover additional experiences... Becker's theory of the household, or theory of addiction , or theory of suicide, or theory of discrimination is not a new theory : it is Chicago-style economic applied to a new subject » (LAWSON, 2004, p.5).

2.2.2.2 Le modèle d'addiction rationnelle à l'épreuve

Dans une spécification économétrique, des facteurs explicatifs inobservables qui entrent dans e_t se retrouvent absorbés dans le terme accidentel. Leur intervention à différentes périodes de consommation rend les variables C_{t-1} et C_{t+1} endogènes. De plus, BECKER (1996) souligne que ces facteurs liés au cycle de vie sont vraisemblablement corrélés en série, élément aggravant l'endogénéité de ces variables dynamiques. Même si l'addiction est inexistante, la corrélation de ces facteurs inobservables ou bien la simple présence d'un terme d'hétérogénéité inobservable invariant dans le temps pour chaque agent suffit à faire dépendre significativement la consommation courante des consommations passée et future. Il s'agit donc d'être très attentif au traitement des variables endogènes et aux méthodes d'estimation employées : recadrées dans le modèle d'addiction rationnelle, les études empiriques à base de modèles dynamiques ne tenant pas compte de cette endogénéité seraient immédiatement invalidées.

Dans la littérature, si la base des spécifications économétriques pour tester le type d'addiction à un bien reste l'équation (2.21), elles sont très souvent élargies à d'autres variables explicatives. BECKER et al. (1994) incluent d'autres facteurs courants tels que des variables de revenu, de taxe, et d'indices de contrebande ; BALTAGI et GRIFFIN (2002) introduisent une variable de revenu et de prix minimum dans un état voisin pour tester l'addiction à l'alcool aux Etats-Unis ; TIEZZI (2004) tient compte du volume consommé de tabac importé de l'étranger pour modéliser le comportement des fumeurs italiens, GARDES et STARZEC (2002) accompagnent leur spécification de variables d'âge, de localisation, d'éducation (parmi d'autres encore) pour tester l'addiction à la consommation d'alcool et de tabac en Pologne. Cependant, ces derniers auteurs remarquent que les régresseurs additionnels devraient entrer dans la fonction de consommation au même titre que les variables e , c'est-à-dire en faisant intervenir simultanément la valeur courante e_t et future e_{t+1} dans la spécification du modèle. Or, la grande majorité des études ne fait pas apparaître cette contrainte et le travail empirique qui suit s'en affranchira également. Ce choix est ici motivé par les possibilités d'exploitation des résultats : en n'intervenant sur la consommation que de manière courante, on peut appliquer à ces variables additionnelles les mêmes outils d'analyse que ceux afférents à la variable (courante) de prix.

Le test de la théorie d'addiction rationnelle a été conduit sur des sources de données très variées. Les premiers travaux empiriques mettant à l'épreuve les modèles d'addiction se sont principalement concentrés sur la consommation de drogues pour des considérations de santé publique, et parce que les modèles d'addiction sont *a priori* les plus enclins actuellement à décrire ce type de consommation. Ainsi, CHALOUPIKA (1991), BECKER (1994), GARDES et STARZEC (2002), et TIEZZI (2004) ont notamment exploité les modèles rationnels pour expliquer la consommation de tabac ; GROSSMAN et al. (1995), BENTZEN et al. (1999), LALLA et al. (2004) les ont appliqués pour modéliser la consommation d'alcool ; VAN OURS (1995) les utilise pour décrire la consommation d'opium ; GROSSMAN et al. (1998) les mettent à l'œuvre pour modéliser la consommation de cocaïne.

Mais, comme le prétendent BECKER et MURPHY (1988), les agents peuvent rendre compte d'une addiction pour un bien quelconque sans pour autant que sa consommation engendre une dépendance biologique ou pharmaceutique.

DEWENTER (2002) reporte même en première page de ses travaux (relatifs à l'addiction aux informations en Allemagne) la citation des auteurs de référence :

*« People get addicted not only to alcohol, cocaine, and cigarettes, but also to work, eating, music television, their standard of living, other people, religion, and many other activities »*²⁸, BECKER et MURPHY (1988).

Ainsi, MOBILIA (1993) a voulu vérifier l'existence d'une addiction au jeu, CAWLEY (1999) utilise la formulation (2.21) pour décrire la consommation énergétique de calories, VILLANI (1992) applique ce modèle sur une consommation d'art, CAMERON (1999) ainsi que SISTO et al. (2005) les utilisent pour vérifier l'hypothèse d'addiction à la consommation de cinéma. A notre connaissance, le modèle (2.21) n'a pas encore été appliqué pour décrire l'usage de la voiture, à l'heure où la « dépendance automobile » est une thématique majeure de la recherche sur les transports.

Les résultats empiriques des travaux qui testent le modèle d'addiction rationnelle restent toutefois généralement mitigés. Si, bien souvent, les effets des consommations passée et future sur le niveau courant de consommation sont significatifs, le taux de préférence intertemporel qui en découle est parfois peu raisonnable. Concernant la consommation de tabac par exemple, BECKER (1996) reporte des taux très élevés allant de 56% à 223%, et TIEZZI (2004) n'obtient pas de valeurs plausibles. Pour la consommation d'alcool, BALTAGI et GRIFFIN (2002), ainsi que GROSSMAN et al. (1998) reportent même quelques taux négatifs. Ces derniers suggèrent de valider l'hypothèse d'addiction rationnelle sur la base d'un taux de préférence convenable, plutôt que prématurément sur la significativité des coefficients associés aux valeurs passée et future de la consommation. AULD et GROONTENDORST (2001) freinent cependant l'enthousiasme du recours au modèle de BECKER et al. (1994), en montrant qu'il peut conclure au comportement d'addiction rationnelle pour la consommation de biens *a priori* non addictifs, tels que les oranges ou les œufs, alors que les résidus d'estimation des modèles ne présentent pas de bonnes propriétés. De plus, ces derniers auteurs mettent en garde contre les résultats du modèle lorsque les données traitées sont des séries temporelles

²⁸ Même la recherche en économie aurait potentiellement un pouvoir addictif selon DEWENTER (2002). S'il est avéré, l'écriture d'une thèse constitue probablement le stock initial de capital addictif justifiant en partie l'activité ultérieure des chercheurs.

agrégées. Plus généralement, LAUX (2000) suggère que ce taux de préférence devrait correspondre à celui qui accompagne les comportements d'épargne ou les décisions d'investissement des agents : l'auteur interprète la différence comme le résultat potentiel d'une rationalité limitée des agents. BALTAGI et GRIFFIN (2001) relient l'inconsistance des taux de préférence au type de données manipulées : ils estiment que l'usage de données agrégées permet rarement de révéler un taux crédible et précis.

2.3 Modèles économétriques et élasticités

2.3.1 Elasticités et politique économique

L'élasticité, dont on doit la paternité à Marshall (DEATON et MUELLBAUER, 1980, p.61), s'est très vite installée comme un outil incontournable pour l'interprétation des résultats des modèles d'analyse de la demande : ses mesures sont effectivement séduisantes tant elles sont facilement interprétables, dénuées de dimension ou d'échelle, et peuvent être dérivées d'une grande majorité des modèles économétriques, même parfois comparées d'une étude à une autre. Dans le domaine de la demande de transport, l'estimation des élasticités à la variation de ses déterminants est une information particulièrement appréciée des pouvoirs publics. En effet, elle permet (outre la simple prévision chiffrée liée à l'évolution de la conjoncture) la conduite de politiques pragmatiques visant à infléchir ou accélérer la consommation d'un bien, selon les objectifs visés. Les mécanismes généralement préférés pour (ré)orienter la consommation sont de nature économique, tels que les variations de taxation et/ou de redistribution des revenus. Ces outils sont plus rapidement mis en place et fonctionnels que d'autres résultant de politiques (par exemple sociales, spatiales...) de plus long terme. Pour cette raison, notre étude privilégie la mesure des effets - prix et des effets - revenu sur l'évolution de l'automobilité des ménages français, mais ses conclusions s'attarderont également sur l'effet de facteurs déterminants plus structurels.

2.3.2 Mesures du modèle statique

La fonction de demande estimée est dite statique lorsque la consommation courante n'est expliquée que par des variables contemporaines. Ce modèle suppose

un ajustement total et instantané (ou au cours de la période observée) de la consommation aux variations de ses déterminants. Ainsi, il suppose absent tout effet temporel dynamique entre des périodes d'observation successives. De fait, c'est le modèle très souvent employé lorsque les données manipulées sont de type transversal. En désignant C_i la quantité consommée, E_i les variables économiques (prix et revenu), S_i l'ensemble des autres déterminants exogènes, et ε_i le terme accidentel, la fonction de demande statique s'écrit :

$$C_i = S_i\alpha_0 + E_i\alpha_1 + \varepsilon_i. \quad (2.29)$$

L'ajustement instantané à l'équilibre lors d'une variation de E_i implique l'égalité des élasticités de la demande (au prix et au revenu) de court et de long terme, respectivement notés $e_{(C/E)}^{CT}$ et $e_{(C/E)}^{LT}$. Calculées à la moyenne de l'échantillon, nous obtenons :

$$e_{(C/E)}^{CT} = e_{(C/E)}^{LT} = \alpha_1 \frac{\bar{E}_i}{\bar{C}_i}. \quad (2.30)$$

2.3.3 Mesures du modèle d'addiction myope

Le modèle retardé est le plus naturellement utilisé lorsque le modélisateur dispose d'une série chronologique, d'un panel ou d'un pseudo-panel et qu'il souhaite séparer facilement les effets de court et de long terme. Par exemple, DARGAY (2004) présente un modèle dynamique expliquant le (log-) kilométrage hebdomadaire par la variable endogène retardée et une série de variables socio-économiques, à partir d'un pseudo-panel construit sur les vagues annuelles du UK Family Expenditure Survey de 1970 à 1995. Il s'agit du modèle à ajustement courant partiel suivant. Notons C_{it}^* la quantité optimale désirée de consommation à la période t définie par $C_{it}^* = S_{it}\alpha_0^* + E_{it}\alpha_1^* + \varepsilon_{it}^*$, alors le niveau de consommation courante dans ce modèle s'écrit :

$$C_{it} = (1 - \theta)C_{it}^* + \theta C_{it-1}, \text{ pour } 0 \leq \theta < 1. \quad (2.31)$$

A l'équilibre, nous avons l'égalité $E(C_{it}) = E(C_{it-1}) = E(C_{it}^*)$. Hors équilibre, nous avons pour $\theta \neq 0$:

$$\begin{aligned} [E(C_{it}) - E(C_{it-1})]/(1 - \theta) &= [E(C_{it}^*) - E(C_{it-1})], \\ [E(C_{it}) - E(C_{it-1})] > 0 &\Rightarrow E(C_{it}) < E(C_{it}^*). \end{aligned} \quad (2.32)$$

Un changement permanent des déterminants économiques induit donc la sortie de l'équilibre et l'ajustement instantané qui suit n'est que partiel pour $\theta \neq 0$. Le retour à un nouvel équilibre ne pourra finalement être atteint qu'après plusieurs périodes. La substitution de C_{it}^* par son expression donne la spécification dynamique usuelle du modèle retardé :

$$C_{it} = \theta C_{it-1} + S_{it}\alpha_0 + E_{it}\alpha_1 + \varepsilon_{it}. \quad (2.33)$$

La relation qui lie les multiplicateurs d'impact instantanés (α_0, α_1) avec les multiplicateurs de long terme (α_0^*, α_1^*) est donnée par :

$$(\alpha_0, \alpha_1) = (1 - \theta) \times (\alpha_0^*, \alpha_1^*). \quad (2.34)$$

On remarque aisément que le modèle statique précédent est un cas particulier du modèle retardé pour une valeur de θ nulle. Plus généralement, le coefficient θ reflète la vitesse d'ajustement des comportements vers le nouvel équilibre : plus il est proche de zéro et plus ils sont rapides ; au contraire, un coefficient proche de 1 reflète un ajustement total très long.

A la moyenne de la population, on dérive les élasticités de court $e_{(C/E)}^{CT}$ et long terme $e_{(C/E)}^{LT}$ par les formules classiques :

$$e_{(C/E)}^{CT} = \alpha_1 \frac{\bar{E}_{it}}{\bar{C}_{it}}, \quad (2.35)$$

$$e_{(C/E)}^{LT} = \alpha_1^* \frac{\bar{E}_{it}}{\bar{C}_{it}} = \frac{\alpha_1}{(1 - \theta)} \frac{\bar{E}_{it}}{\bar{C}_{it}}. \quad (2.36)$$

Pour $0 < \theta < 1$, alors les valeurs d'élasticité précédentes respectent l'inégalité $|e_{(C/E)}^{CT}| < |e_{(C/E)}^{LT}|$.

2.3.4 Mesures du modèle d'addiction rationnelle

Une version de l'équation de demande du modèle d'addiction rationnelle en (2.21), augmentée de variables explicatives courantes est donnée par :

$$C_{it} = \theta C_{it-1} + \frac{\theta}{1 + \rho} C_{it+1} + S_{it} \alpha_0 + E_{it} \alpha_1 + \varepsilon_{it}.$$

Le prix P_t du bien addictif C est inclus dans le vecteur des variables économiques E_{it} , alors que S_{it} rassemble des facteurs explicatifs d'autre nature.

BECKER et al. (1994, p.113) déclinent les expressions de l'effet sur la consommation actuelle C_{it} induit par une variation occasionnelle ou permanente (à différents horizons) des variables explicatives exogènes, par exemple E_{it} . A l'aide des racines caractéristiques φ_1 et φ_2 calculées en (2.27) et (2.28), les valeurs d'élasticités de la demande du bien C aux variations des déterminants E , évaluées aux moyennes \bar{E}_t et \bar{C}_t , sont données par les formules suivantes :

- E1 : Elasticité de C_{it} à une variation occasionnelle non anticipée de E_{it} :

$$\alpha_1 (\theta \varphi_2)^{-1} \times (\bar{E}_t / \bar{C}_t). \quad (2.37)$$

- E2 : Elasticité à une variation occasionnelle non anticipée de E_{it-1} :

$$\varphi_2^{-1} \alpha_1 (\theta \varphi_2)^{-1} \times (\bar{E}_t / \bar{C}_t). \quad (2.38)$$

- E3 : Elasticité à une variation occasionnelle non anticipée de E_{it+1} :

$$\varphi_1 \alpha_1 (\theta \varphi_2)^{-1} \times (\bar{E}_t / \bar{C}_t). \quad (2.39)$$

- E4 : Elasticité à une variation occasionnelle anticipée de E_{it} :

$$\frac{\alpha_1}{(\theta(\varphi_2 - \varphi_1))} \times \frac{\bar{E}_t}{\bar{C}_t}. \quad (2.40)$$

- E5 : Elasticité à une variation permanente (immédiate) de E_{it} à court terme :

$$\frac{\alpha_1}{(\theta(1 - \varphi_1)\varphi_2)} \times \frac{\bar{E}_t}{\bar{C}_t}. \quad (2.41)$$

- E6 : Elasticité à une variation permanente (sur toutes les périodes) de E_{it} à long terme :

$$\frac{-\alpha_1}{(\theta(1 - \varphi_1)(1 - \varphi_2))} \times \frac{\bar{E}_t}{\bar{C}_t}. \quad (2.42)$$

Les élasticités (E1), (E2), (E3) expriment la sensibilité moyenne des consommations à une déviation temporaire (lors d'une seule période) et non anticipée des déterminants économiques courants, passés et futurs. L'impact sur la consommation courante d'un changement temporaire anticipé, c'est-à-dire connu de si longue date par les agents qu'ils ont eu le temps d'ajuster leur trajectoire de consommation, est calculé en (E4). L'élasticité de court terme induite par une variation permanente d'un déterminant économique²⁹ (E5) calcule la sensibilité de la consommation en la période initiale du changement, alors que l'élasticité de long terme mesure cette variation après une infinité de périodes (E6).

Enfin, on remarque facilement que les élasticités issues du modèle retardé (ou d'addiction myope) sont des cas particuliers de celles du modèle d'addiction rationnelle pour $\varphi_1 = 0$. En effet, puisque les individus myopes ne tiennent pas compte du futur dans leur choix présent ($B = \theta / (1 + \rho) = 0$), il vient que l'effet anticipé d'un changement économique futur n'a aucun effet sur la consommation courante et que l'élasticité calculée par (E3) est nulle. Pour la même raison, la myopie entraîne l'égalité des élasticités anticipée (E1), non anticipée (E4), et de l'élasticité de court terme d'un changement permanent (E5).

2.4 Données d'application et statistiques descriptives

2.4.1 Les données

Dans le premier chapitre, nous avons présenté l'enquête Parc Auto qui constitue la matière de travail permanente dans ce mémoire. Il n'est toutefois pas

²⁹ Pour cette élasticité, on considère que le changement permanent est immédiat : les consommations passées sont maintenues constantes et contraignent les choix actuels et futurs de consommation.

inutile de rappeler ici comment celle-ci est organisée. La mise en forme annuelle de l'enquête est effectuée en deux fichiers distincts. Le premier fichier, dit « fichier - ménage », regroupe l'ensemble des caractéristiques des ménages répondant, et décrit notamment pour chacun sa tranche de revenus annuels, sa localisation résidentielle, et des caractéristiques de ses membres (âge, sexe, possession de permis, catégorie professionnelle, etc...). Le second fichier, dit « fichier - voiture », fait un descriptif de leur motorisation en détaillant jusque trois automobiles à disposition des ménages répondant. Plus particulièrement, ce fichier renseigne pour chaque véhicule son kilométrage annuel parcouru³⁰, son âge, le carburant utilisé. Un numéro d'identifiant unique et invariable dans le temps pour chaque ménage permet un appariement exact de ces deux types de fichiers d'observations pour une année particulière, ainsi que le suivi longitudinal individuel pour les ménages répondant à plusieurs vagues d'enquête. Le modèle d'addiction rationnelle nécessite l'observation individuelle sur au moins trois périodes successives. Nous avons donc pu le caler sur le panel pondéré Parc Auto 1999-2001, dont le montage a été décrit en section 1.3.

2.4.1.1 Traitement des fichiers de véhicules : l'imputation des variables manquantes

La description déclarative par les ménages du kilométrage annuel de leurs automobiles dans l'enquête Parc Auto peut parfois sembler douteuse. D'autres fois, celui-ci n'est tout simplement pas renseigné. Plus précisément, près d'une automobile sur dix dans chaque vague du panel est décrite sans kilométrage annuel, une sur cent fait état d'un kilométrage annuel suspect inférieur à 100 km, et les répondants pourraient exagérer l'utilisation d'environ une voiture sur cent en 2001 (≥ 50000 km)³¹. L'existence de quelques points aberrants et de non-réponses est une caractéristique quasi inévitable des sources de données détaillées au niveau

³⁰ Si un remplacement de voiture est intervenu en cours d'année, alors le questionnaire de l'enquête suggère d'additionner les kilométrages de l'ancienne et de la nouvelle automobile.

³¹ Pour les années 1999 et 2000, l'organisme collecteur des données (SOFRES) a simplement borné ces kilométrages aberrants à 50000km.

microéconomique³², et leur dissolution dans un ensemble important de bonnes observations n'est pas toujours garantie³³.

Nous avons choisi un mode d'imputation par prédicteurs (ARMOOGUM, 2000) à ces valeurs faiblement crédibles ou inexistantes, ainsi qu'au kilométrage des voitures n'ayant réellement pas roulé de l'année³⁴. Celui-ci utilise une régression (ici linéaire) du kilométrage des bons répondants sur un ensemble de leurs caractéristiques. Sans fondement théorique, nous avons cherché les paramètres d'un modèle très prédictif en coupe instantanée pour décrire le kilométrage des voitures, semble-t-il, convenablement renseigné dans chaque vague annuel du panel. Le modèle retenu, de forme semi-log, est exprimé en fonction du niveau de gamme du véhicule, de son carburant, de sa tranche d'âge³⁵, ainsi que des caractéristiques de l'utilisateur principal et de sa zone de résidence. L'imputation aux valeurs manquantes, aberrantes ou nulles est effectuée par le calcul $km_{it}^{imp} = \exp(X_{it}\hat{\beta}_t + \hat{\varepsilon}_{it})$, avec : $\hat{\beta}_t$ le vecteur des paramètres estimés du modèle utilisant la t -ième vague du panel ; X_{it} le vecteur des caractéristiques du répondant i ; $\hat{\varepsilon}_{it}$ un résidu simulé suivant $\mathcal{N}(0, \hat{\sigma}_t^2)$; $\hat{\sigma}_t^2$ l'erreur quadratique moyenne du t -ième modèle. La forme exponentielle garantit que les kilométrages imputés sont positifs : ceci permet d'obtenir et d'exploiter un parc automobile dit « roulant », mais aussi d'éviter les biais que peuvent générer les comportements d'usage atypiques³³.

Cette méthode pratique d'imputation, même si elle reste critiquable si le mécanisme de sélection entre bons et mauvais répondants n'est pas aléatoire, permet de conserver l'ensemble des observations et de constituer une « *clean data matrix* » opérationnelle, c'est-à-dire sans manquants. Les résultats des modèles pour l'imputation sont présentés dans la table 2.1. La dernière étape pour la constitution des données est la fusion par ménage de leurs caractéristiques avec celles de leurs véhicules lorsqu'ils existent.

³² Surtout, comme pour l'enquête Parc Auto, lorsque les formulaires sont « auto-administrés », c'est-à-dire qu'ils s'appuient sur les seules déclarations des interrogés, sans recours à un enquêteur formé « sur mesure » pour la collecte des informations.

³³ Par exemple, HUBER (1981) a montré que 3% d'observations anormales pouvaient affaiblir considérablement la qualité des estimations d'un modèle.

³⁴ Les « véritables zéros », dont on vérifie la validité à partir d'une question de l'enquête, demandant explicitement si la voiture n'a pas roulé en cours d'année.

³⁵ Dans chaque vague, l'âge des véhicules est obtenu par la différence de l'année d'enquête avec l'année de la première mise en circulation.

TABLE 2.1 : Résultats des modèles servant l'imputation aux valeurs manquantes/aberrantes

Année	1999		2000		2001	
Variable dépendante : ln(kilométrage du véhicule)	Coefficient	(t-ratio)	Coefficient	(t-ratio)	Coefficient	(t-ratio)
constante	9.60	(171.07)	9.50	(153.93)	9.53	(144.81)
Nombre d'autres automobiles dans le ménage	-0.07	(-3.80)	-0.09	(-4.60)	-0.09	(-4.47)
Motorisation du ménage (référence : moteur essence)						
- Moteur diesel	0.42	(15.53)	0.46	(16.11)	0.44	(15.36)
- Moteur non spécifié	0.28	(3.25)	0.36	(2.88)	0.25	(1.88)
Age du véhicule (référence : moins de 5 ans)						
- 5 à 9 ans	-0.15	(-5.13)	-0.17	(-5.55)	-0.13	(-4.09)
- 10 à 14 ans	-0.36	(-10.28)	-0.38	(-10.51)	-0.40	(-10.95)
- Plus de 15 ans	-0.89	(-16.19)	-0.91	(-16.88)	-0.83	(-15.53)
- Age non renseigné	-0.52	(-5.93)	-0.67	(-7.19)	-0.56	(-5.32)
Genre de l'utilisateur principal (référence : homme)						
- Femme	-0.19	(-6.43)	-0.17	(-5.53)	-0.14	(-4.64)
Age de l'utilisateur principal (référence : 18-29 ans)						
- 30-39 ans	-0.07	(-1.39)	-0.03	(-0.51)	-0.09	(-1.55)
- 40-49 ans	-0.06	(-1.25)	-0.05	(-1.00)	-0.04	(-0.66)
- 50-59 ans	-0.22	(-4.37)	-0.13	(-2.47)	-0.13	(-2.33)
- 60-69 ans	-0.28	(-4.25)	-0.28	(-3.65)	-0.34	(-4.44)
- Plus de 70 ans	-0.45	(-6.04)	-0.45	(-5.51)	-0.58	(-6.93)
CSP de l'utilisateur principal (référence : employé)						
- Agriculteurs exploitants	0.11	(0.92)	0.13	(1.04)	-0.12	(-0.87)
- Artisans, commerçants, Chefs d'entreprise	-0.08	(-1.12)	-0.03	(-0.34)	-0.03	(-0.26)
- Cadres et professions Intellectuelles supérieures	0.11	(1.74)	0.11	(1.86)	0.19	(3.16)
- Professions intermédiaires	0.09	(1.99)	0.12	(2.48)	0.08	(1.61)
- Ouvriers	-0.04	(-0.75)	0.01	(0.13)	0.02	(0.49)
- Retraités	-0.08	(-1.31)	0.02	(0.24)	0.06	(0.78)
- Sans activité professionnelle	-0.10	(-1.76)	0.01	(0.15)	-0.15	(-2.66)
- Sans réponse	-0.04	(-0.73)	-0.02	(-0.12)	0.04	(0.40)
Localisation du ménage (référence : Province)						
- Paris	-0.23	(-2.40)	-0.20	(-1.94)	-0.38	(-3.68)
- Petite couronne francilienne	-0.08	(-1.26)	-0.14	(-2.22)	-0.11	(-1.77)
- Grande couronne francilienne	0.05	(1.01)	0.08	(1.57)	-0.06	(-1.11)
Propriétés des modèles :						
Nombre d'observations (de véhicules)	3285		3300		3331	
R ²	>0.99		>0.99		>0.99	
EQM ($\hat{\sigma}_t^2$)	0.49		0.55		0.57	

Note : estimations à partir des automobiles des ménages participants au panel pondéré Parc Automobile 1999-2001.

2.4.2 Statistiques descriptives

La constitution du panel avait permis d'identifier 3010 ménages continûment présents dans les vagues 1999 à 2001 de l'enquête Parc Automobile. La proportion annuelle des non-motorisés est légèrement inférieure à 20%, celle des mono-motorisés se situe globalement à 50%, et celle des bi-motorisés est légèrement supérieure au quart. Les ménages tri-motorisés et davantage représentent un peu moins de 5% de l'échantillon total (table 2.2).

TABLE 2.2 : Taux de motorisation par année (pourcentages)

Année	1999	2000	2001
0 voiture	19.03	18.16	18.49
1 voiture	50.31	50.39	49.61
2 voitures	25.85	26.83	27.00
3 voitures et plus	4.80	4.62	4.91

Source : panel Parc Automobile 1999-2001.

Ces proportions masquent des inégalités spatiales très fortes : plus de 60% des ménages parisiens sont non-équipés en automobile, soit une part environ deux fois plus élevée qu'en petite couronne parisienne (PC), et quatre fois plus importante qu'en grande couronne parisienne (GC) ou en Province. Ainsi, on compte en moyenne près de deux fois plus de voitures par ménage en petite couronne francilienne qu'à Paris, et près de trois fois plus en Province (table 2.3)

TABLE 2.3 : Nombre moyen d'automobiles par ménage par zone

Année	Paris	PC	GC	Province	France
1999	0.43	0.87	1.16	1.24	1.18
2000	0.46	0.83	1.18	1.26	1.19
2001	0.45	0.85	1.14	1.27	1.20

Source : panel Parc Automobile 1999-2001.

On dénombre 3667 automobiles décrites par les ménages panélistes en 1999, 3686 en 2000 et 3664 en 2001. Pour ces années respectives, l'âge moyen des véhicules est de 6.37, 6.77, et 6.76 ans. En fonction du carburant, nous observons que les véhicules essence sont plus âgés que les voitures diesel d'environ deux années (table 2.4).

TABLE 2.4 : Age moyen du parc par type de motorisation

Année	1999		2000		2001	
Age (en années)	Moy.	Ec-type	Moy.	Ec-type	Moy.	Ec-type
Véhicule diesel	5.19	4.01	5.52	4.28	5.48	4.38
Véhicule essence	7.22	5.70	7.50	5.86	7.56	5.99
Imprécis	7.08	5.79	6.53	5.02	6.17	4.71

Source : panel Parc Automobile 1999-2001.

Dans nos données, l'automobilité moyenne de l'ensemble des ménages accuse une diminution continue entre 1999 et 2001, passant de 15610 km à 14826 km. Similairement, la mobilité moyenne des ménages équipés évolue de 19279 km en 1999 à 18188 km en 2001, soit un retrait moyen de 1100 km environ. Là encore, de fortes disparités spatiales sont observables : celles-ci relèvent en partie des différences de taux de motorisation, mais aussi du type d'équipement automobile possédé. En 2000 par exemple, l'automobilité moyenne des ménages équipés parisiens et de petite couronne était d'environ 14500 km, celle des ménages franciliens de grande couronne était d'environ 17700 km, alors que celle des ménages provinciaux était de 19000 km. (table 2.5). Globalement, on observe une répartition du parc automobile entre voitures diesel et essence équivalentes en Île-de-France quelle que soit la zone : $\frac{1}{4}$ des automobiles sont à moteur diesel alors que $\frac{3}{4}$ fonctionnent aux supercarburants. Toutefois, les voitures diesel sont nettement plus représentées en Province avec une proportion d'environ $\frac{1}{3}$ de son parc. Au regard de nos données, le processus de diésélisation du parc français³⁶ décrit par HIVERT (1996a) semble plutôt prendre sa source en Province qu'en Île-de-France,

³⁶ Dans notre panel, la part du diesel dans le parc total croît de 2 points par an (table 2.6).

où cette évolution de fond est ici moins franche (table 2.6). Ce phénomène est notamment explicatif de l'évolution de la consommation énergétique moyenne des automobiles puisque les voitures diesel sont plus économes que les voitures essence d'environ 0.9 litres de carburant pour cent kilomètres parcourus (table 2.7), ou encore, d'environ 90 grammes-équivalent-pétrole³⁷. On soupçonne également une amélioration temporelle de l'efficacité pour ces deux types de véhicule au long des trois années : globalement dans nos données, une automobile consommait en moyenne 7.33 litres de carburant pour cent kilomètres parcourus en 2001 contre 7.44 initialement en 1999.

TABLE 2.5 : Automobilité des ménages équipés par zone de résidence
(en kilomètres)

Année	1999		2000		2001	
Zone	Moyenne	Ec-type	Moyenne	Ec-type	Moyenne	Ec-type
Paris	13594	9305	14498	9013	12301	10177
PC	15396	9649	14792	9924	14753	9024
GC	18377	11218	17656	10288	16408	10351
Province	19719	13033	18969	12611	18677	12889
France	19279	12759	18563	12312	18188	12575

Note : parc roulant (après imputations).

Source : panel Parc Automobile pondéré 1999-2001.

³⁷ Soit environ 1.6% de « mieux » que la consommation énergétique des voitures essence en gep (grammes-équivalent pétrole) en 2001. Un litre de gazole équivaut à 840 gep, contre 750 gep pour un litre de supercarburant (HIVERT et al., 2006, p.42).

TABLE 2.6 : Répartition dans le parc des types de motorisation par carburant et par zone (pourcentages)

Année	Type	Paris	PC	GC	Province	France
1999	diesel	24	27	27	35	34
	essence	75	72	71	63	64
	imprécis	1	1	2	2	2
2000	diesel	24	26	26	37	36
	essence	76	74	73	61	63
	imprécis	0	0	1	1	1
2001	diesel	25	24	28	39	38
	essence	75	74	71	60	61
	imprécis	0	2	1	1	1

Source : panel Parc Automobile 1999-2001.

TABLE 2.7 : Evolution de la consommation de carburant par type de véhicule (en litres de carburant pour 100 km parcourus)

Année	1999		2000		2001	
Efficacité	Moyenne	Ec-type	Moyenne	Ec-type	Moyenne	Ec-type
Véhicule diesel	6.82	1.29	6.80	1.49	6.74	1.35
Véhicule essence	7.74	1.63	7.72	1.58	7.67	1.56
Imprécis	8.52	3.02	8.14	2.90	8.55	1.52
Parc total	7.44	1.62	7.39	1.66	7.33	1.55

Source : panel Parc Automobile 1999-2001.

Le prix annuel moyen au litre des carburants a globalement augmenté entre 1999 et 2001³⁸, avec un record en 2000 à la suite de l'envolée atypique du cours du baril de brut. Cette année là, un litre de gazole coûtait en moyenne 5.55 F, alors que le prix moyen à la pompe pour les supercarburants s'élevait à 7.47 F (table 2.8).

³⁸Rappelons qu'avant la mise en circulation de l'Euro (€) en France au début de l'année 2002, les prix étaient libellés en Francs (F).

TABLE 2.8 : Evolution des prix moyens à la pompe des carburants en France par année et par type de motorisation (en Francs par litre)

Année	1999	2000	2001
Gazole	4.520	5.550	5.220
Supercarburants	6.445	7.465	7.135
Imprécis	5.803	6.827	6.497

Notes : le prix du litre de carburant pour les voitures imprécises est calculé par une moyenne pondérée des prix au litre du gazole et des supers. Le prix des supercarburants est une moyenne arithmétique des prix moyens annuels des supers sans plomb et du super.
Source : Calculs INRETS (2003), à partir du Comité Professionnel Du Pétrole (CPDP), et de l'Observatoire De l'Energie.

Ces différences de prix et d'efficacité expliquent en partie l'utilisation moyenne nettement plus intensive des voitures diesel. Ainsi, une automobile diesel a parcouru en moyenne 17085 km en 2000 contre 10412 km pour une voiture essence. Entre 1999 et 2001, la tendance générale est à la baisse de ces kilométrages moyens, les voitures diesel parcourant environ 1600 km de moins, contre 840 km pour les voitures essence (table 2.9).

TABLE 2.9 : Kilométrage annuel moyen des automobiles par type de motorisation

Année	1999		2000		2001	
Kilométrage (en km)	Moyenne	Ec-type	Moyenne	Ec-type	Moyenne	Ec-type
Diesel	17786	9014	17085	9099	16187	8662
Essence	11001	6564	10412	6138	10157	6172
Imprécis et autres	13274	8273	13913	8947	13063	8208
Parc total	13381	8184	12899	8038	12477	7826

Note : parc roulant (après imputations). Source : panel Parc Automobile 1999-2001.

Enfin, l'ensemble des moyennes pour les autres variables d'intérêt et relatives aux ménages pour cette étude sont synthétisées dans la table 2.10. La table

2.11, qui reporte les consommations moyennes de carburants et les kilométrages moyens en 2000 par automobile³⁹ selon le double critère du type de motorisation et de la zone résidentielle du ménage, servira le calcul des élasticités au prix des carburants dans une prochaine section.

TABLE 2.10 : Statistiques descriptives des caractéristiques des ménages

Année	1999		2000		2001	
Variable	Moyenne	Ec-type	Moyenne	Ec-type	Moyenne	Ec-type
Nombre d'automobiles	1.18	0.82	1.19	0.80	1.20	0.82
Nombre d'adultes	1.86	0.74	1.86	0.74	1.87	0.75
Nombre d'actifs occupés	1.00	0.85	0.98	0.85	0.98	0.86
Nombre de femmes adultes	0.98	0.48	0.98	0.48	0.98	0.48
Nombre de mineurs	0.52	0.93	0.50	0.91	0.49	0.90
Nombre de permis	1.55	0.82	1.56	0.81	1.57	0.81
Revenu du ménage (kF)	151.38	89.75	156.24	91.51	162.06	94.27
Chef du ménage âgé de :						
moins de 40 ans (%)	0.34	0.47	0.31	0.46	0.29	0.45
40 à 60 ans (%)	0.42	0.49	0.42	0.49	0.42	0.49
plus de 60 ans (%)	0.24	0.43	0.27	0.44	0.29	0.45
Répartition résidentielle :						
Paris (%)	0.05	0.22	0.05	0.22	0.05	0.22
Petite Couronne Francilienne (%)	0.06	0.24	0.06	0.24	0.06	0.24
Grande Couronne Francilienne (%)	0.07	0.25	0.07	0.25	0.07	0.25
Province (%)	0.82	0.38	0.82	0.38	0.82	0.38
Automobilité (km) (ménages motorisés uniquement)	19279	12759	18563	12313	18189	12576
Automobilité (km) (tous les ménages)	15610	13752	15193	13240	14826	13371

Source : panel Parc Automobile 1999-2001 (après imputations).

³⁹ Exceptés les kilométrages imputés, rappelons que les variables du kilométrage annuel et de la consommation des voitures sont renseignées et de façon uniquement déclarative.

TABLE 2.11 : Kilométrages moyens et consommations moyennes des automobiles, selon le type de véhicule et la zone de résidence en 2000

Zone géographique	Type de carburant	Kilométrage annuel moyen des automobiles (en km)	Consommation moyenne des automobiles (en litres/100 km)
Province	Gazole	17087	6.79
	Supers	10358	7.65
Grande couronne	Gazole	17041	6.73
	Supers	11421	7.84
Petite couronne	Gazole	16278	7.07
	Supers	10020	8.30
Paris	Gazole	19086	6.89
	Supers	9726	8.27
France entière	Gazole	17085	6.80
	Supers	10412	7.72

Note : parc roulant (après imputations).

Source : panel pondéré Parc Automobile 1999-2001.

2.5 Modélisation et difficultés

2.5.1 Détail du modèle utilisé

Nous utilisons les notations suivantes pour formaliser notre modèle :

- les indices i, v, t identifiant respectivement les ménages, les automobiles, et les périodes, $i \in [1; 3010]$, $v \in [1; 3]$, $t \in [1999; 2001]$,

- les variables KM , CSM , PD , PE , PI renvoyant respectivement au kilométrage annuel, à la consommation moyenne de carburant pour cent kilomètres, au prix du litre de gazole, au prix du litre de supercarburant et à une moyenne pondérée de ces prix pour les voitures imprécises.

- la variable PKM , produit de l'efficacité énergétique du véhicule et du prix du carburant utilisé, calculant ainsi le prix individuel pour la consommation de cent kilomètres : $PKM_{ivt} = CSM_{ivt} \times PC_{ivt}$, avec PC renseignant le prix du carburant utilisé (PD ou PE ou PI).

- les variables AVD , AVE , AVI , NRA renseignant respectivement l'âge du véhicule s'il possède une motorisation diesel, essence ou imprécise, et une indicatrice si l'âge n'est pas renseigné.

Dans une première étape, le kilométrage annuel d'un véhicule d'un ménage équipé est modélisé en utilisant la spécification linéaire suivante :

$$KM_{it} = \beta_{0it} + \beta_1 PKM_{it} + \beta_2 AVD_{it} + \beta_3 AVE_{it} + \beta_4 AVI_{it} + \beta_5 NRA_{it} + \varepsilon_{it}, \quad (2.43)$$

avec : $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2), \quad (2.44)$

où la constante du modèle β_{0it} est ici différenciée selon que le ménage est domicilié en t à Paris, en petite couronne francilienne, en grande couronne francilienne, ou en Province. L'automobilité du ménage est obtenue par la sommation des kilométrages des voitures qu'il détient. Le modèle, ainsi agrégé au niveau du ménage, devient :

$$KM_{it} = \sum_v KM_{it} = \beta_0 NV_{it} + \beta_1 \sum_v PKM_{it} + \beta_2 \sum_v AVD_{it} + \beta_3 \sum_v AVE_{it} + \beta_4 \sum_v AVI_{it} + \beta_5 \sum_v NRA_{it} + \sum_v \varepsilon_{it}, \quad (2.45)$$

avec NV_{it} , la variable du nombre de voitures en t du ménage différenciée selon sa localisation géographique. Afin de mieux tenir compte de l'hétérogénéité des ménages, la spécification linéaire précédente a été augmentée par l'introduction de variables socio-éco-démographiques des ménages suivantes (libellés entre parenthèses) :

- la valeur médiane de la tranche de revenu (REV),
- trois indicatrices d'âge du chef de famille : entre 18 et 39 ans (JU), entre 40 et 65 ans (AM), et plus de 65 ans (SE),
- le nombre d'adultes autres que le chef de ménage (NAD),
- le nombre d'actifs occupés (NAC),
- le nombre de femmes adultes (NFE),
- le nombre de permis de conduire (NP),
- le nombre de mineurs (NMI).

Enfin, le modèle inclut également comme déterminants la valeur du kilométrage annuel passé (KM_{t-1}) et futur (KM_{t+1}) selon le type de modèle d'addiction (myope ou rationnel) mis en application (section 2.2.2).

2.5.2 Biais de sélection

Les modèles visités jusqu'ici nécessitent que la variable dépendante ne soit pas censurée. Or les ménages non-motorisés déclarent une automobilité nulle, qui correspond en microéconomie à une solution dite «en coin». Les modèles classiques prenant en compte les solutions en coin des agents rendent compte d'une structure souvent non linéaire, comme les modèles Tobit (TOBIN, 1958) et de CRAGG (1971). Or, puisque nous voulons conserver la forme linéaire initiale, il nous a fallu sélectionner dans l'échantillon les ménages n'ayant aucune solution en coin pour l'année centrale d'estimation, en $t = 2000$. Le biais de sélection potentiel qui en résulte est corrigé par l'introduction de l'inverse d'un ratio de Mills, noté λ , parmi l'ensemble des régresseurs. Ce type de correction est assez usuel lorsque l'estimation d'une forme économétrique linéaire ne s'effectue que sur une catégorie ciblée d'observations (SIMMONET et ULRICH, 2001). Il s'agit d'une méthode de correction initialement proposée par HECKMAN (1979) qui repose sur le traitement d'une corrélation potentielle entre le terme d'erreur du modèle de sélection et celui du modèle de demande.

Avant son introduction dans le modèle principal (2.45), l'inverse du ratio de Mills est préalablement estimé en utilisant les résultats d'un modèle dichotomique probit, qui évalue pour chaque ménage la probabilité d'être équipé en l'an 2000, à partir des variables explicatives les plus descriptives du niveau d'équipement. Nous utilisons celles du nombre de permis de conduire, de zone résidentielle, et de revenu, dont les poids très significatifs sur le niveau d'équipement sont confirmés dans le troisième chapitre de ce mémoire. La table 2.12 reproduit les résultats d'estimation du modèle de sélection de première étape, dont est issu une estimation $\hat{\lambda}_i$ de λ_i , pour chaque ménage. L'introduction de la variable $\hat{\lambda}$ parmi les facteurs explicatif constitue par ailleurs une source d'hétéroscédasticité puisqu'elle conditionne la variance du terme résiduel de (2.45) aux régresseurs utilisés dans le modèle de sélection (HECKMAN, 1979)⁴⁰. Enfin, le test de la significativité du coefficient estimé associé à $\hat{\lambda}$ constitue un test d'existence du biais de sélection.

⁴⁰ Bien que HECKMAN (1986) nous donne les éléments de sa correction, nous privilégierons d'autres méthodes plus pratiques et globales puisqu'il existe d'autres facteurs d'hétéroscédasticité, identifiés dans le paragraphe suivant. Par ailleurs, l'usage comme régresseur de variables estimées telles que $\hat{\lambda}$ dans une spécification a pour effet de sous-estimer les écarts-types des paramètres si leur effet est significatif : dans ce cas, une correction s'impose. Si la variable $\hat{\lambda}$ est non significative, la correction peut être ignorée.

TABLE 2.12 : Résultats du modèle de sélection

Variables	Coefficient	t-ratio
Indicatrices de la tranche d'âge du chef de famille (Référence : entre 40 et 64 ans)		
Entre 18 et 39 ans	-0.101	-0.800
Plus de 65 ans	0.223	2.190
Indicatrices de résidence du ménage (Référence : Ménage parisien)		
Ménage en petite couronne	0.677	3.810
Ménage en grande couronne	1.643	8.100
Ménage en Province	1.652	11.130
Nombre d'adultes	0.070	0.910
Nombre d'actifs occupés	0.112	1.430
Nombre de femmes	-0.363	-4.050
Nombre de mineurs	0.145	2.440
Revenu annuel (en kF)	0.217	4.910
Nombre de permis	1.503	20.220
Constante	-2.625	-12.990

Notes : vague 2000 du panel Parc Automobile 1999-2001, 3010 ménages. Estimation probit pondéré, variable dépendante : $Y_i = 1$ si le ménage est motorisé ; $Y_i = 0$ si non-motorisé.

Log vraisemblance : -747.71309, pseudo R^2 : 0.4734.

2.5.3 Hétéroscédasticité

Les ménages dans notre panel n'ont pas la même importance, celle-ci est caractérisée par l'affectation d'un poids pour chacun, noté w_i , que nous avons déterminé par la méthode des scores de propension. Cette pondération des observations peut être considérée sous la forme d'un schéma d'hétéroscédasticité connu de la variance des aléas quand la spécification économétrique est linéaire. Lorsque l'on dispose du vecteur des pondérations, il est pratique d'en tenir simplement compte en transformant les données selon la méthode d'AITKEN (1935), c'est-à-dire en multipliant les variables individuelles par $\sqrt{w_i}$.

Cependant, nous pressentons d'autres facteurs d'hétéroscédasticité. L'hypothèse assumant que les aléas ε_{ivt} entre voitures sont indépendants et homoscédastiques pour les ménages multi-motorisés est, semble-t-il, peu crédible. En effet, un choc négatif sur l'aléa d'une voiture d'un ménage telle une panne nécessitant son immobilisation temporaire peut induire un regain d'usage (un choc positif) sur l'aléa d'autres voitures du ménage : typiquement, cela désigne une

corrélation des aléas : $\text{cov}(\varepsilon_{ivt}, \varepsilon_{iv't}) \neq 0$. Par ailleurs, les chocs aléatoires induits par des conditions particulières de circulation ou d'offre de transport public peuvent être différemment répercutés sur le kilométrage des voitures d'un ménage, mais aussi sur le kilométrage des automobiles selon le degré de motorisation du ménage ou sa zone de résidence : $\text{var}(\varepsilon_{ivt}) \neq \text{var}(\varepsilon_{i'v't})$. La construction du modèle lors de la sommation du kilométrage des véhicules en (2.45) est également productrice d'hétéroscédasticité, liée au niveau de motorisation des ménages. Ainsi, nous devons appliquer des méthodes d'estimation qui prennent en compte l'existence d'une forme complexe d'hétéroscédasticité.

2.5.4 Endogénéité

En impliquant simultanément les variables dépendantes retardée et avancée, la spécification dynamique particulière du modèle d'addiction rationnelle rend ces déterminants nécessairement endogènes, même en admettant l'hypothèse d'indépendance temporelle des erreurs individuelles⁴¹. De plus, BECKER (1996) insiste sur le fait que les erreurs sont vraisemblablement corrélées en série, par exemple à cause d'un effet d'hétérogénéité individuel inobservable η_i constant dans le temps : $\sum_v \varepsilon_{ivt} = \varepsilon_{it} = \eta_i + u_{it}$, qui vient renforcer la plausibilité d'une corrélation entre $KM_{it\pm 1}$ et ε_{it} . Dans ces circonstances, l'usage des moindres carrés ordinaires (MCO) conduit irrémédiablement à une estimation biaisée des paramètres, ce qui nous amène à considérer d'autres méthodes d'ajustement pour les modèles dynamiques.

2.6 Méthodes d'estimation et tests

Cette section détaille les outils statistiques théoriques ayant servi notre traitement empirique des données. Leur présentation s'inspire du document méthodologique de BAUM, SHAFFER et STILLMAN (2003). En présence de variables explicatives endogènes dans une spécification, l'estimateur des MCO fournit des estimations biaisées des paramètres. L'usage courant dans cette situation est le

⁴¹ Pour s'en rendre compte, il suffit d'écrire la formulation du modèle pour $(t-1)$ et $(t+1)$, et d'établir que les covariances entre le terme résiduel en t et ces deux variables endogènes décalées sont différentes de zéro lorsque les coefficients associés sont non nuls.

recours aux estimateurs à variables instrumentales (VI). L'estimateur naturel qui se présente est celui des doubles moindres carrés ordinaires (2MCO). Cependant, si cet estimateur fournit des estimations convergentes des paramètres, il reste inefficace lorsque la variance des aléas des observations est hétéroscédastique, et l'inférence statistique est impossible. Il est dès lors nécessaire de trouver une méthode d'estimation dite « robuste ».

Si nous avons pu corriger celle liée à la pondération des observations, nous avons vu que coexistaient vraisemblablement d'autres motifs d'hétéroscédasticité. Pratiquement, il convient de s'en assurer, en testant cette hypothèse via la procédure bien connue de BREUSH et PAGAN (1979). Si la présence d'hétéroscédasticité des aléas est validée, alors la méthode (instrumentale et robuste) d'estimation par les moments généralisés (MMG) doit être adaptée⁴². Cet outil, proposé par HANSEN (1982), généralise d'autres estimateurs tels que ceux des MCO et 2MCO, et est de fait très apprécié des chercheurs appliqués. La condition de sa mise en œuvre réside dans la disposition de bons instruments, c'est-à-dire assez corrélés avec les variables explicatives endogènes et indépendants du résidu du modèle. La vérification de ces deux critères devra être examinée, tout d'abord par un test d'exogénéité des instruments (HANSEN, 1982), puis par le test de significativité des variables instrumentales sur les variables endogènes (BOUND, 1995 ; SHEA, 1997).

2.6.1 L'estimateur de la méthode des moments généralisés

Lorsque les erreurs sont hétéroscédastiques, le meilleur estimateur à base de variables instrumentales est donné par celui de la méthode des moments généralisés⁴³, noté MMG. Cette méthode repose sur la recherche des conditions d'orthogonalité entre le terme d'erreur du modèle et l'ensemble des variables exogènes (GREENE, chap.18, 2003). Après les avoir transformés par la méthode d'AITKEN (1935) pour tenir compte des poids individuels w_i , notons Y la variable dépendante, X l'ensemble des variables explicatives, Z l'ensemble des variables exogènes, ε le vecteur des erreurs du modèle linéaire $Y = X\beta + \varepsilon$, puis notons

⁴² Dans la pratique, l'hypothèse H_0 d'homoscédasticité des aléas doit être testée après une estimation par 2MCO, en utilisant la statistique de Breush-Pagan. Si l'hypothèse est acceptée, alors l'estimateur MMG donne des résultats similaires à ceux obtenus par 2MCO. Au contraire, son rejet justifie ainsi le recours à l'estimateur MMG.

⁴³ En anglais, « Generalized Method of Moment » (GMM).

$W = \sum_i w_i$ le poids total de l'échantillon. L'orthogonalité des instruments avec les erreurs doit satisfaire la relation théorique suivante :

$$E(Z_i' \varepsilon_i) = E(Z_i'(Y_i - X_i \beta)) = 0. \quad (2.46)$$

En cherchant la correspondance de (2.46) dans les données, l'objectif est de trouver le vecteur de paramètres $\hat{\beta}$ qui annule le vecteur des moments empiriques suivant :

$$\frac{1}{W} Z'(Y - X\hat{\beta}) = 0. \quad (2.47)$$

Lorsque l'on dispose de plus de variables instrumentales dans Z que de paramètres à estimer dans β , le modèle est dit « suridentifié » : on ne peut pas généralement trouver de solution unique pour β . Dans ce cas, on utilise une matrice de pondération P carrée de format $l \times l$, où l est le nombre de variables instrumentales, puis on utilise une forme quadratique des moments empiriques précédents pour former le critère q à minimiser :

$$q(\hat{\beta}) = W \left[\frac{e(X\hat{\beta})' Z}{W} \right] P \left[\frac{Z' e(X\hat{\beta})}{W} \right] = W \bar{m}(\hat{\beta})' P \bar{m}(\hat{\beta}), \quad (2.48)$$

avec :

$$\bar{m}(\hat{\beta}) = \frac{1}{W} Z' e(\hat{\beta}),$$

$e(\hat{\beta}) = (Y - X\hat{\beta})$, l'erreur d'observation,

P , la matrice de pondération MMG.

Le critère q peut être interprété comme une fonction de perte ou de distance. La résolution du problème de minimisation donne l'estimateur des moments généralisés :

$$\hat{\beta}_{GMM} = (X' Z P Z' X)^{-1} (X' Z P Z' Y). \quad (2.49)$$

Celui-ci généralise beaucoup d'autres estimateurs linéaires traditionnels. Notamment, en choisissant $Z = X$ et $P = W(X'X)^{-1}$, on retrouve l'estimateur

des moindres carrés ordinaires (MCO) ; ou encore l'estimateur des doubles moindres carrés (2MCO) en prenant comme matrice de pondération $P = W(Z'Z)^{-1}$. Il y a cependant autant d'estimateurs MMG qu'il peut exister de matrices P , mais l'estimateur MMG le plus efficace est celui possédant la matrice de pondération optimale, qui minimise la variance asymptotique de l'estimateur MMG. Celle-ci est donnée en choisissant pour P la matrice S^{-1} , avec :

$$S = \frac{1}{W}E\left(Z'\varepsilon \varepsilon'Z\right) = \frac{1}{W}(Z'\Omega Z). \quad (2.50)$$

Pour $P = S^{-1}$, alors nous avons $Wq(\hat{\beta}_{GMM}) \sim \mathcal{X}^2$, et l'estimateur (2.49) devient un estimateur de Khi carré minimal. Pour que l'estimateur MMG efficace puisse être calculable, nous devons avoir en main une estimation consistante de S , et donc une estimation de Ω . WHITE (1980) a montré qu'un estimateur consistant (pas nécessairement efficace) de β était suffisant pour estimer des matrices de variances - covariances consistantes d'estimateurs. L'estimateur 2MCO, qui possède cette propriété, permet d'obtenir des résidus $e(\hat{\beta}_{2MCO})$ qui sont également de consistants estimateurs de ε . Ainsi, en l'absence de corrélation des aléas entre ménages, un estimateur consistant de Ω est donné par :

$$\hat{\Omega} = \text{diag}\left(e\left(\hat{\beta}_{2MCO}\right)^2\right), \quad (2.51)$$

et un estimateur consistant de S et donné par $\hat{S} = W^{-1}(Z'\hat{\Omega}Z)$. On dispose maintenant de tous les éléments permettant de calculer l'estimateur MMG efficace, ainsi que sa matrice de variance - covariance :

$$\hat{\beta}_{MMGE} = \left(X'Z(\hat{S}^{-1})Z'X\right)^{-1}\left(X'Z(\hat{S}^{-1})Z'Y\right), \quad (2.52)$$

$$\text{Asym.Var}\left(\hat{\beta}_{MMGE}\right) = W \times \left(X'Z(\hat{S}^{-1})Z'X\right)^{-1}. \quad (2.53)$$

Relativement à l'estimateur 2MCO, la considération de l'hétéroscédasticité en (2.51) ne justifie pas le nom « triple moindres carrés » pour $\hat{\beta}_{MMGE}$, déjà réservé par ailleurs à un estimateur de modèles à équations simultanées. Cherchant une appellation au reflet de sa construction, DAVIDSON et MACKINNON (1993) ont par

exemple proposé « double moindre carré hétéroscédastiques⁴⁴ », alors que CUMBY et al. (1983) avaient avancé celui de « double moindres carrés à deux étapes⁴⁵ ». Il existe d'autres estimateurs MMG praticables : par exemple, le vecteur des résidus du modèle estimé par la méthode MMG efficace en (2.52) peut à son tour être récupéré pour calculer une nouvelle matrice \hat{S} , qui est ensuite utilisée pour recalculer une nouvelle estimation MMG efficace en (2.52)..., et ainsi de suite jusqu'à la stabilisation des estimations obtenues. Cette procédure d'estimation est connue sous le nom d'« *estimateur MMG itératif* ».

2.6.2 Le test de Breush et Pagan

L'homoscédasticité des aléas dans une régression est l'une des hypothèses pour la mise en application des estimateurs MCO et à variables instrumentales. Considérée à tort, c'est-à-dire en présence d'hétéroscédasticité des erreurs entre les observations, la convergence de ces estimateurs n'est pas affectée, mais leur matrice des covariances est rendue inconsistante, et donc non utilisable pour l'inférence. BREUSH et PAGAN (1979) ont posé la base d'un test post-estimateur fondé sur la corrélation entre le carré du terme résiduel et une série de variables explicatives. Considérons le modèle initial : $Y_i = X_i\beta + \varepsilon_i$, avec $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ par hypothèse H_0 d'homoscédasticité. Une estimation $\hat{\beta}$ de β par MCO ou 2MCO est utilisée pour calculer le vecteur des termes résiduels associé, $\hat{\varepsilon} = Y - X\hat{\beta}$, à partir duquel on obtient une estimation usuelle de la variance des aléas $\hat{\sigma}^2$, et les moments suivants : $E(\hat{\varepsilon}_i^2) = \sigma^2$, $V(\hat{\varepsilon}_i^2) = 2\sigma^4$.

Sous l'hypothèse H_0 d'homoscédasticité des aléas, alors $E(\hat{\varepsilon}_i^2)$ doit être constant quelle que soit l'observation i , et donc ne pas dépendre d'un ensemble de variables noté \tilde{Z}_i . Le test de Breush et Pagan est un test de la non-significativité de ces variables à expliquer le carré des résidus empiriques $\hat{\varepsilon}_i^2$, par l'ajustement du modèle suivant : $\hat{\varepsilon}_i^2 = \sigma^2(1 + \tilde{Z}_i\alpha + u_i)$. En notant $\tilde{\varepsilon}_i^2 = (\hat{\varepsilon}_i^2 / \sigma^2 - 1)$, le modèle devient : $\tilde{\varepsilon}_i^2 = \tilde{Z}_i\alpha + u_i$, avec $E(u_i) = 0$, $V(u_i) = 2$. En remplaçant σ^2 par son estimation convergente sous H_0 $\hat{\sigma}^2$, l'estimateur MCO de α ($\hat{\alpha}$) nous est alors donné par : $\hat{\alpha} = (\tilde{Z}'\tilde{Z})^{-1} \tilde{Z}'\tilde{\varepsilon}^2$. Nous avons ainsi $(\hat{\alpha} - \alpha) \sim \mathcal{N}(0, 2(\tilde{Z}'\tilde{Z})^{-1})$. On calcule la statistique BP de Breush-Pagan par :

⁴⁴ De sigle H2SLS, pour « Heteroskedastic two Stage Least Squares ».

⁴⁵ De sigle 2S2SLS, pour « Two-Step Two Stage Least Squares ».

$$BP = \frac{(\hat{\alpha}'(\tilde{Z}'\tilde{Z})\hat{\alpha})}{2}. \quad (2.54)$$

Sous l'hypothèse H_0 d'homoscédasticité, alors $BP \sim \mathcal{X}^2(p)$, avec p : le nombre de variables dans la matrice \tilde{Z} . Si la valeur empirique de la statistique BP excède la valeur théorique pour une erreur de première espèce donnée, on rejette l'hypothèse H_0 : il est donc nécessaire de calculer une nouvelle matrice de covariances de l'estimateur (MCO ou 2MCO), telle que la matrice « sandwich » de Eicker-Huber-White (WHITE, 1980 ; BAUM et al., 2003), ou d'utiliser un estimateur de β qui soit robuste à la présence d'hétéroscédasticité des aléas ε_i , tel que l'estimateur MMG.

2.6.3 Le test de Hansen

Les estimateurs à base de variables instrumentales reposent sur l'orthogonalité présumée des instruments avec le processus d'erreur du modèle. Lorsqu'il y a plus d'instruments exclus de la spécification que de variables instrumentées, on peut tester cette hypothèse en calculant la statistique J de Hansen (HANSEN, 1982) :

$$J(\hat{\beta}_{MMGE}) = Wq(\hat{\beta}_{MMGE}) = e(\hat{\beta}_{MMGE})'Z(Z'\hat{\Omega}Z)^{-1}Z'e(\hat{\beta}_{MMGE}) \quad (2.55)$$

Sous l'hypothèse H_0 de l'orthogonalité des instruments utilisés avec le vecteur des erreurs, alors $J(\hat{\beta}_{MMGE}) \sim \chi^2_{l-k}$, avec l le nombre total d'instruments (inclus et exclus du modèle), et k le nombre total de régresseurs dans le modèle. Si la statistique calculée $J(\hat{\beta}_{MMGE})$ excède la valeur théorique pour une p-valeur à 5%, ou similairement, si la p-valeur de la distribution théorique calculée en $J(\hat{\beta}_{MMGE})$ est inférieure à 5%, alors l'hypothèse H_0 est rejetée, au profit de l'hypothèse alternative H_1 d'une corrélation entre les termes d'erreurs et le lot des instruments Z . Si le test de Hansen conduit à rejeter l'hypothèse d'indépendance, alors deux motifs peuvent être invoqués : des instruments sont exclus à tort de la régression (cas de variables omises), ou bien certains instruments ne sont pas purement exogènes (mauvais instruments). Le test de Hansen généralise celui de Sargan (SARGAN, 1956) à la présence d'hétéroscédasticité des erreurs. En cas d'homoscédasticité, ils sont équivalents.

2.6.4 Le test de Bound

Outre leur condition d'orthogonalité avec le processus d'erreur, de bons instruments exclus du modèle doivent également avoir un pouvoir explicatif statistiquement significatif sur les régresseurs endogènes défaillants. Notons X_D l'ensemble des régresseurs « défaillants » du modèle qui sont endogènes, Z_1 l'ensemble des instruments exogènes exclus de la spécification du modèle (en réserve), et Z_2 l'ensemble des instruments exogènes inclus dans la spécification du modèle.

Expliquer X_D par les seuls instruments Z_2 sans que les instruments Z_1 n'aient de véritable pouvoir explicatif revient à instrumenter X_D par une combinaison linéaire des variables Z_2 : se pose alors pour la spécification du modèle un problème d'une colinéarité stricte entre régresseurs. Le test de Bound (BOUND, 1995) propose donc de juger de la pertinence des instruments Z_1 à expliquer davantage que Z_2 les régresseurs endogènes X_D . Dans sa forme la plus simple, il teste la significativité jointe des instruments exclus Z_1 dans chaque régression robuste des variables endogènes de X_D sur l'ensemble des instruments $Z = \{Z_1; Z_2\}$. Dans l'usage, le plus simple est d'effectuer un F-test de nullité des paramètres associés aux instruments Z_1 . Après chaque régression des variables contenues dans X_D sur Z , notons, $\hat{\delta}_1$ le sous vecteur des paramètres estimés associé à Z_1 , $Var(\hat{\delta}_1)$ la sous-matrice des covariances des paramètres associée à Z_1 , l_1 et l , respectivement le nombre d'instruments dans Z_1 et dans Z . Sous l'hypothèse H_0 que $\delta_1 = 0$, alors la statistique de Bound, notée F_{Bound} , suit une distribution de Fisher :

$$F_{Bound} = \frac{(\hat{\delta}_1' Z_1' Z_1 \hat{\delta}_1) / l_1}{\left[\begin{array}{c} \text{somme des carrés des} \\ \text{résidus du modèle} \\ X_D = Z\delta + e_D \end{array} \right] / (N - l)} \sim F[l_1, N - l]. \quad (2.56)$$

Si la p-valeur calculée en F_{Bound} dans la distribution théorique $F[l_1, N - l]$ est inférieure à 5%, alors l'hypothèse H_0 est réfutée : les instruments exclus Z_1 ont donc un pouvoir explicatif additionnel à celui des instruments Z_2 sur les régresseurs endogènes X_D . Ils constituent donc un réservoir d'instruments opérationnel s'ils sont également exogènes.

Néanmoins, le défaut du test réside dans sa recherche de significativité *jointe* : il suffit en effet d'un seul bon instrument explicatif de l'ensemble des régresseurs endogènes pour que l'hypothèse H1 soit systématiquement validée, alors même que les autres instruments exclus peuvent n'avoir aucun effet explicatif sur les variables X_D . Si le nombre des bons instruments est insuffisant par rapport au nombre de régresseurs endogènes à instrumenter, alors un modèle ajusté par variables instrumentales devient sous-identifié, donc impossible à estimer. Mais s'il n'y a qu'une variable à instrumenter, un test de Bound qui accepte l'hypothèse alternative suffit pour valider l'usage des instruments contenus dans Z_1 . Lorsqu'il y a plusieurs régresseurs endogènes, on a donc besoin d'autres indicateurs pour valider le groupe d'instruments Z_1 .

SHEA (1997) propose l'usage d'une statistique sous la forme d'un R^2 partiel permettant de surmonter cette défaillance possible du test de Bound. En comparant sa statistique avec le R^2 partiel de Bound, les règles d'interprétation lorsque l'hypothèse de significativité est antérieurement acceptée par le F-test de Bound sont les suivantes⁴⁶ :

- si le R^2 de Bound est fort par rapport au R^2 de Shea, alors il n'y a pas assez d'instruments efficaces contenus dans Z_1 pour suffisamment expliquer les régresseurs endogènes ;
- si les deux R^2 sont de valeurs proches (ils sont égaux lorsqu'il n'y a qu'une variable à instrumenter), alors Z_1 constitue un bon réservoir d'instruments explicatifs des régresseurs endogènes X_D .

2.7 Résultats

Les résultats d'estimation des modèles de demande visités dans ce chapitre sont reportés en table 2.13. Nous allons tout d'abord faire l'examen de leurs propriétés générales avant d'aborder l'interprétation économique, les mesures d'élasticité aux prix et au revenu, puis les effets marginaux induits par une modification exogène de déterminants.

⁴⁶ Voir GODFREY(1999) pour le détail technique du R^2 partiel de Shea.

2.7.1 Qualité des modèles et tests

Tous les modèles présentent une bonne qualité d'ajustement, avec un coefficient de détermination (R^2) compris entre 0.84 et 0.90. Ainsi, le F-test de significativité jointe de l'ensemble des régresseurs utilisés dans chacun confirme leur pouvoir explicatif sur l'automobilité des ménages. Pour les modèles estimés par MCO, la statistique de Breush-Pagan conduit à rejeter l'hypothèse nulle d'homoscédasticité des aléas entre ménages. La correction des écarts-types d'estimation est effectuée à l'aide de la matrice « sandwich » de Eicker-White (WHITE, 1980). La statistique de Breush-Pagan suivant une estimation des modèles rationnels par 2MCO rejette également l'hypothèse d'homoscédasticité, justifiant ainsi le recours à l'estimateur robuste MMG comme technique de variable instrumentale.

Pour les modèles d'addiction, les instruments utilisés dans la mise en application de l'estimateur MMG sont l'ensemble des variables exogènes du modèle estimé (instruments inclus) et l'ensemble des caractéristiques des ménages augmenté des indicatrices de leur localisation, à la période passée et future. Le test de Hansen⁴⁷ conduit à accepter, dans les deux modèles d'addiction estimés par MMG, l'hypothèse nulle d'orthogonalité entre les résidus d'estimation et les instruments. Par ailleurs, les tests de Bound et de Shea acceptent l'hypothèse alternative de significativité jointe des instruments exclus sur les variables explicatives endogènes lorsque l'effet des instruments inclus est « neutralisé ». Les instruments exclus possèdent donc les bonnes propriétés pour être utilisés : ils sont significativement influents sur les variables endogènes⁴⁸, tout en restant indépendants du résidu d'estimation. L'usage des prix passés et futurs pour instrumenter les consommations passées et futures des agents, possibilité suggérée par BECKER (1996), avait pour effet de dégrader sérieusement les tests d'indépendance de Hansen : ils n'ont donc pas été utilisés.

⁴⁷ Dans leur très bon tutoriel sur les méthodes à base de variables instrumentales, BAUM et al. (2003) estiment que « *this test ... should be performed as a standard diagnostic in any overidentified instrumental variables estimation* », en accord avec l'opinion de Sargan (reprise dans ce tutoriel), pour qui présenter des résultats dans ce contexte sans tester l'hypothèse d'indépendance est une « fraude pieuse ».

⁴⁸ Mais peut être pas suffisamment : STAIGER et STOCK (1997) (cités par BAUM et al., 2003) ont par notamment montré que de faibles instruments peuvent aussi conduire à une significativité du F-test de Bound.

2.7.2 Biais de sélection ?

Dans un premier temps, tous les modèles s'accordent à rejeter l'hypothèse de significativité du terme de correction λ , les t-ratio présentés étant largement inférieur à 1.96. Il n'existerait donc pas de déterminants inobservables régissant à la fois la décision d'équipement et l'intensité d'usage des automobiles.

2.7.3 Addiction et taux de préférence intertemporel

Alors que les estimations MCO et MMG semblent accrédi-ter l'hypothèse d'un comportement addictif de l'usage de l'automobile aux dépens du modèle statique, c'est le modèle d'addiction rationnelle qui semble le plus pertinent pour décrire le type de dépendance des ménages à la voiture.

Quelle que soit la méthode utilisée en effet, les modèles myopes confirment l'influence significative du kilométrage passé pour déterminer le kilométrage courant des ménages, validant ainsi l'hypothèse d'une dynamique temporelle de l'usage de l'automobile.

Les modèles rationnels rejettent unanimement la myopie des comportements par une significativité du coefficient de la variable endogène avancée : selon l'interprétation de BECKER et al. (1994), les ménages auraient donc conscience des effets futurs de leur consommation actuelle et en tiendraient compte dans leurs choix courants. Cependant, l'estimation du modèle MCO conduit à un taux de préférence intertemporel peu plausible de -22%. Celles non biaisées issues du modèle MMG révèlent un taux beaucoup plus crédible de 17% : à cet égard, notre étude privilégie l'exploitation des résultats de ce modèle.

TABLE 2.13 : Estimations des modèles statiques et dynamiques de la demande de kilomètres (addiction myope et rationnelle).
Variable dépendante : automobilité du ménage en 2000 ($KM_{t=2000}$)

Variable	i - Statique (MCO robuste)	ii - Addiction myope (MCO robuste)	iii - Addiction rationnelle (MCO robuste)	iv - Addiction myope (MMG)	v - Addiction rationnelle (MMG)
KM_{t-1}	-	0.452 (13.85)	0.305 (9.45)	0.307 (3.41)	0.346 (4.20)
KM_{t+1}	-	-	0.391 (12.73)	-	0.295 (3.98)
REV	17.79 (4.23)	10.23 (2.70)	4.52 (1.43)	13.32 (3.98)	7.07 (2.16)
$NV_{(\text{province})}$	16648.04 (14.08)	8262.89 (7.98)	5116.39 (5.53)	11038.71 (5.56)	6189.06 (2.96)
$NV_{(\text{IDF-GC})}$	16150.22 (12.27)	7992.72 (6.87)	5203.99 (4.94)	10810.79 (5.50)	6310.44 (3.08)
$NV_{(\text{IDF-PC})}$	13782.60 (9.93)	6391.53 (4.83)	4194.43 (3.46)	8540.45 (4.46)	4888.76 (2.63)
$NV_{(\text{IDF-Paris})}$	12932.08 (7.22)	6750.86 (4.19)	4655.87 (3.25)	8813.36 (4.45)	5514.74 (2.94)
$\sum_v PKM_v$	-86.59 (-4.57)	-37.98 (-2.06)	-20.82 (-1.22)	-55.29 (-3.04)	-34.04 (-1.98)
$\sum_v AVD_v$	-54.82 (-0.65)	18.17 (0.21)	36.07 (0.40)	-24.75 (-0.34)	-12.39 (-0.18)
$\sum_v AVE_v$	-293.90 (-6.87)	-149.21 (-4.17)	-82.82 (-2.59)	-191.68 (-4.04)	-104.054 (-2.28)
$\sum_v AVI_v$	-176.05 (-1.34)	-55.83 (-0.49)	-105.92 (-0.95)	-57.04 (-0.46)	-30.2560 (-0.28)
$\sum_v NRA_v$	-3920.08 (-5.62)	-1727.34 (-2.86)	-990.12 (-1.88)	-2439.82 (-3.20)	-1112.06 (-1.53)
NP	2686.86 (3.09)	1653.09 (2.14)	1238.23 (1.99)	2138.45 (3.13)	1324.65 (2.16)
NAD	1156.27 (1.98)	1082.01 (1.98)	810.40 (1.71)	510.27 (1.01)	409.90 (0.96)

Variable (suite)	Statique (MCO robuste)	Addition myope (MCO robuste)	Addition rationnelle (MCO robuste)	Addition myope (MMG)	Addition rationnelle (MMG)
<i>NAC</i>	1227.27 (3.39)	275.10 (0.79)	-5.04 (-0.02)	815.36 (2.07)	286.32 (0.76)
<i>NFE</i>	-2668.04 (-3.89)	-1477.33 (-2.56)	-954.96 (-2.06)	-2041.94 (-3.34)	-1158.26 (-2.16)
<i>NMI</i>	-322.17 (-1.00)	-357.72 (-1.14)	-470.60 (-1.59)	-174.637 (-0.61)	-297.28 (-1.16)
<i>AM</i>	-1634.10 (-2.36)	-1921.60 (-2.94)	-1628.90 (-2.61)	-1617.14 (-2.64)	-1376.77 (-2.51)
<i>SE</i>	-3644.43 (-4.60)	-2966.92 (-3.92)	-1805.55 (-2.54)	-2648.16 (-3.68)	-1608.01 (-2.34)
$\hat{\lambda}$	2751.41 (1.43)	1784.89 (1.03)	1487.93 (1.06)	2430.34 (1.71)	1555.84 (1.26)
constante	-1820.72 (-0.97)	-1225.91 (-0.74)	-1459.53 (-1.10)	-1210.78 (-0.90)	-954.01 (-0.85)
ρ	—	∞	-22.02%	∞	17.04%
R^2	0.84	0.88	0.90	0.87	0.90
Fisher (p-valeur)	F(18, 2612) : 66.6 (0.00)	F(19, 2611) : 94.5 (0.00)	F(20, 2610) : 133.8 (0.00)	F(19, 2611) : 82.6 (0.00)	F(20, 2610) : 115.5 (0.00)
Breuch-Pagan (p-valeur)	χ^2 (18) : 653.5 (0.00)	χ^2 (19) : 865.4 (0.00)	χ^2 (20) : 788.5 (0.00)	χ^2 (40) : 777.0 (0.00)	χ^2 (40) : 878.8 (0.00)
Hansen (p-valeur)	—	—	—	χ^2 (21) : 30.46 (0.08)	χ^2 (20) : 23.22 (0.28)
Bound (p-valeur)	—	—	—	F(22, 2590) KMt-1 : 3.80 (0.00)	F(22, 2590) KMt-1 : 3.80 (0.00) KMt+1 : 4.51 (0.00)

Notes : 2631 observations, 11 instruments exclus pour 1999, 11 pour 2001, 19 instruments inclus dans la régression, dont une constante et le lambda d'Heckman. Les estimateurs sont robustes à l'hétéroscédasticité. T-ratios des paramètres entre parenthèses. La statistique de Breush-Pagan est calculée à partir des résidus d'estimation de méthodes inefficaces (MCO et 2MCO) pour évaluer la pertinence de l'usage d'estimateurs robustes. Les valeurs des R^2 de Bound et de Shea (non reportées, mais disponibles sur demande) sont très proches pour le modèle d'addiction rationnelle estimé par MMG (colonne v).

2.7.4 Elasticités prix

Même si les coefficients estimés des prix pour 100 kilomètres de carburant consommés sont de signe attendus, négatifs, leur significativité diffère entre les modèles. L'effet prix est particulièrement significatif pour les modèles statique et d'addiction myope quelle que soit la méthode d'estimation, alors que celui du modèle d'addiction rationnelle estimé par MCO présente un faible t-ratio qui dégrade sa probabilité d'être en deçà de zéro à 89%. Toutefois, ce dernier modèle estimé par MMG présente un coefficient prix significatif au seuil d'erreur à 5%, malgré un t-ratio légèrement inférieur à 2 (-1.98). Les sensibilités estimées du kilométrage au prix, à court et long terme, et pour l'ensemble des modèles sont reportées dans la table 2.14. On remarque globalement que les élasticités des modèles estimés par MCO sont inférieures à celles obtenues par MMG.

TABLE 2.14 : Elasticités au prix du kilomètre de court et long terme

Elasticités – prix	Statique (MCO)	Addiction myope (MCO)	Addiction rationnelle (MCO)	Addiction myope (MMG)	Addiction rationnelle (MMG)
Court terme	-0.34 [-0.46;-0.22]	-0.14 [-0.27;-0.03]	-0.17 [-0.44; 0.05]	-0.22 [-0.34;-0.10]	-0.23 [-0.41;-0.05]
Long terme	-0.34 [-0.46;-0.22]	-0.27 [-0.50;-0.06]	-0.27 [-0.71; 0.07]	-0.31 [-0.47; -0.16]	-0.37 [-0.72;-0.08]

Notes : élasticités évaluées au ménage moyen, à partir des estimations de la table 2.13.

Intervalles de confiance à 95% calculés par bootstrap paramétrique.

La table 2.15 détaille pour chaque zone géographique l'ensemble des sensibilités au prix du kilomètre à partir du modèle d'addiction rationnelle estimé par MMG : l'élasticité de court terme d'une augmentation permanente s'échelonne d'environ -0.23 en Province et en grande couronne, jusque -0.27 à Paris et en petite couronne francilienne. A long terme, les élasticités respectives sont estimées à -0.38 et à -0.44. Pour une variation accidentelle et non anticipée du prix courant, les valeurs d'élasticité pour ces zones sont respectivement d'environ -0.15 et -0.18 (-0.18

et -0.20 si elle est anticipée). Les élasticités aux variations du prix passé et du prix futur s'établissent environ à -0.06 pour toutes les zones.

TABLE 2.15 : Elasticités au prix du kilomètre par zone de résidence

Elasticités - prix	France entière	Province	GC	PC	Paris
Occasionnel					
passé non anticipé	-0.059 [-0.070;-0.046]	-0.058 [-0.068;-0.047]	-0.063 [-0.065;-0.050]	-0.072 [-0.076;-0.061]	-0.070 [-0.077;-0.056]
courant, non anticipé	-0.150 [-0.264;-0.026]	-0.148 [-0.268;-0.031]	-0.161 [-0.288;-0.029]	-0.183 [-0.330;-0.028]	-0.178 [-0.322;-0.034]
futur non anticipé	-0.050 [-0.055;-0.040]	-0.049 [-0.051;-0.040]	-0.054 [-0.056;-0.043]	-0.061 [-0.066;-0.049]	-0.059 [-0.063;-0.049]
anticipé	-0.173 [-0.297;-0.033]	-0.170 [-0.301;-0.038]	-0.185 [-0.324;-0.037]	-0.211 [-0.369;-0.035]	-0.205 [-0.359;-0.041]
Permanent					
court terme	-0.225 [-0.415;-0.050]	-0.222 [-0.402;-0.040]	-0.242 [-0.434;-0.048]	-0.275 [-0.502;-0.047]	-0.268 [-0.486;-0.055]
long terme	-0.370 [-0.718;-0.099]	-0.365 [-0.696;-0.094]	-0.397 [-0.759;-0.090]	-0.451 [-0.854;-0.086]	-0.439 [-0.861;-0.094]

Notes : élasticités évaluées au ménage moyen de chaque zone, à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13. Intervalles de confiance à 95% calculés par bootstrap paramétrique. PC (GC) pour petite (grande) couronne francilienne.

En niveau, une augmentation permanente de 1 Franc du prix pour effectuer cent kilomètres s'accompagne d'une diminution du kilométrage annuel estimée à 58 km par automobile dans le court terme, et à 95 km à long terme (table 2.20). La table 2.11 qui reporte les consommations moyennes de carburant et les kilométrages moyens par type de véhicule (essence ou diesel) et par zone permet de mesurer l'effet d'une variation du prix des carburants à la pompe sur le kilométrage des automobiles, et d'en dériver des élasticités⁴⁹ du kilométrage automobile aux prix du gazole et de l'essence (table 2.16).

⁴⁹ Exemple de lecture : la consommation moyenne de carburant pour un véhicule diesel en 2000 était de 6.80 litres de gazole pour cent kilomètres parcourus. L'augmentation de 1 Franc du prix du gazole entraîne donc une augmentation de 6.80 Francs pour effectuer

TABLE 2.16 : Elasticités au prix des carburants par zone géographique.

Elasticités– prix	Carbu- rant	France entière	Province	GC	PC	Paris
Occasionnel						
passé non anticipé	<i>gazole</i>	-0.033 [-0.036;-0.028]	-0.033 [-0.035;-0.027]	-0.033 [-0.035;-0.027]	-0.036 [-0.038;-0.030]	-0.030 [-0.032;-0.025]
	<i>supers</i>	-0.083 [-0.087;-0.068]	-0.083 [-0.088;-0.067]	-0.077 [-0.083;-0.064]	-0.093 [-0.098;-0.076]	-0.095 [-0.101;-0.078]
courant non anticipé	<i>gazole</i>	-0.085 [-0.096;-0.076]	-0.085 [-0.095;-0.076]	-0.084 [-0.095;-0.076]	-0.093 [-0.104;-0.084]	-0.077 [-0.086;-0.068]
	<i>supers</i>	-0.213 [-0.236;-0.193]	-0.212 [-0.236;-0.191]	-0.197 [-0.220;-0.179]	-0.238 [-0.265;-0.217]	-0.244 [-0.272;-0.221]
futur non anticipé	<i>gazole</i>	-0.028 [-0.030;-0.024]	-0.028 [-0.030;-0.023]	-0.028 [-0.030;-0.024]	-0.031 [-0.033;-0.025]	-0.026 [-0.027;-0.021]
	<i>supers</i>	-0.071 [-0.075;-0.059]	-0.071 [-0.075;-0.058]	-0.066 [-0.070;-0.055]	-0.079 [-0.084;-0.065]	-0.081 [-0.087;-0.067]
anticipé	<i>gazole</i>	-0.098 [-0.125;-0.084]	-0.098 [-0.125;-0.084]	-0.097 [-0.123;-0.084]	-0.107 [-0.137;-0.092]	-0.089 [-0.115;-0.076]
	<i>supers</i>	-0.245 [-0.314;-0.210]	-0.244 [-0.310;-0.210]	-0.227 [-0.292;-0.197]	-0.273 [-0.353;-0.237]	-0.281 [-0.365;-0.242]
Permanent						
court terme	<i>gazole</i>	-0.128 [-0.188;-0.100]	-0.127 [-0.185;-0.100]	-0.127 [-0.184;-0.099]	-0.139 [-0.203;-0.109]	-0.116 [-0.169;-0.091]
	<i>supers</i>	-0.319 [-0.461;-0.251]	-0.318 [-0.463;-0.252]	-0.296 [-0.435;-0.232]	-0.357 [-0.530;-0.279]	-0.366 [-0.546;-0.289]
long terme	<i>gazole</i>	-0.209 [-0.404;-0.144]	-0.209 [-0.415;-0.142]	-0.208 [-0.389;-0.143]	-0.228 [-0.439;-0.156]	-0.190 [-0.372;-0.129]
	<i>supers</i>	-0.524 [-1.029;-0.358]	-0.522 [-1.028;-0.357]	-0.485 [-0.956;-0.338]	-0.586 [-1.154;-0.400]	-0.601 [-1.192;-0.413]

Notes : élasticités évaluées aux automobiles moyennes de chaque zone en 2000 (table 2.11), à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13, et du prix des carburants de la table 2.8 (voir note 49). Intervalles de confiance à 95% calculés par bootstrap paramétrique.

cette distance. La même année, le kilométrage moyen pour ce type de véhicule est de 17085 kilomètres, et le prix du litre de gazole est de 5.55 francs. De plus, l'effet marginal de long terme du prix est estimé à -95 kilomètres. L'élasticité de long terme du kilométrage au prix de cent kilomètres résulte finalement du calcul : $(-95 \times 6.80) \times (5.55 / 17085) = -0.209$.

Il en résulte assez logiquement que les élasticités du kilométrage annuel au prix du gazole sont toujours de plus faible amplitude que celles au prix des supercarburants, le rapport des sensibilités se situant globalement autour de 2.5. Plus en détail, la sensibilité de court terme à une augmentation permanente du prix des supercarburants est estimée à -0.31 pour les voitures des ménages provinciaux et celles de grande couronne francilienne, et à -0.36 pour les automobiles parisiennes et de petite couronne. Pour ces deux zones, les sensibilités à long terme sont respectivement de -0.50 et -0.59. Concernant l'usage du parc des voitures diesel, l'élasticité au prix du gazole sur leur kilométrage tombe à -0.13 environ à court terme quelle que soit la zone de résidence, et se situe dans le long terme à environ -0.21 pour la grande couronne francilienne et la Province, à -0.19 à Paris, et -0.23 en petite couronne d'Île-de-France. Concernant les changements temporaires de prix, les sensibilités à la variation non anticipée du prix courant du gazole se situent globalement autour de -0.08, alors que celles relatives au prix des supercarburants s'échelonnent de -0.20 en grande couronne francilienne jusque -0.24 à Paris et en petite couronne. Lorsque la variation de prix est anticipée, les élasticités au prix du gazole sont proches de -0.10 sur l'ensemble des zones, et les élasticités au prix des supercarburants varient de -0.28 pour une voiture essence à Paris à -0.23 en grande couronne. Enfin, les sensibilités aux variations de prix passés et futurs non anticipées sont relativement stables quelle que soit la zone géographique : autour de -0.03 pour le gazole et de -0.08 pour les supercarburants.

2.7.5 Elasticités au revenu

Les coefficients estimés associés à la variable de revenu du ménage présentent des caractéristiques similaires à ceux des prix : ils sont globalement significatifs et de signe attendu (positif). C'est particulièrement le cas pour l'ensemble des modèles d'addiction myope et statique. Dans le modèle d'addiction rationnelle, la significativité du coefficient tombe à 85% lorsqu'il est estimé par MCO, mais reste bien établie à 98% lorsqu'il est estimé par MMG. Pour ce dernier modèle, l'effet *ceteris paribus* d'une augmentation permanente du revenu annuel de 1000 F est une augmentation significative du kilométrage du ménage de 12 km à court terme et de 20 km à long terme (table 2.20). Pour chaque modèle, la table 2.17 reporte les élasticités de l'automobilité des ménages à leur revenu. Globalement

sur l'ensemble des modèles, celles ci s'établissent aux environs de +0.10 à court terme, et aux alentours de +0.15 à long terme.

TABLE 2.17 : Elasticités au revenu de court et long terme des modèles

Elasticités – revenu	Statique (MCO)	Addiction myope (MCO)	Addiction rationnelle (MCO)	Addiction myope (MMG)	Addiction rationnelle (MMG)
Court terme	0.15 [0.09 ; 0.21]	0.09 [0.04 ; 0.14]	0.08 [-0.01 ; 0.18]	0.11 [0.06 ; 0.16]	0.10 [0.03 ; 0.16]
Long terme	0.15 [0.09 ; 0.21]	0.16 [0.06 ; 0.26]	0.13 [-0.02 ; 0.29]	0.16 [0.09 ; 0.23]	0.16 [0.06 ; 0.29]

Notes : élasticités évaluées au ménage moyen, à partir des estimations de la table 2.13.

Intervalles de confiance à 95% calculés par bootstrap paramétrique.

TABLE 2.18 : Elasticités au revenu par zone de résidence

Elasticités– revenu	France entière	Province	GC	PC	Paris
Occasionnel					
passé non anticipé	0.026 [0.021 ; 0.027]	0.024 [0.019 ; 0.026]	0.035 [0.028 ; 0.036]	0.043 [0.034 ; 0.044]	0.048 [0.039 ; 0.050]
courant, non anticipé	0.067 [0.020 ; 0.111]	0.061 [0.017 ; 0.104]	0.089 [0.020 ; 0.151]	0.110 [0.029 ; 0.184]	0.123 [0.030 ; 0.208]
futur non anticipé	0.022 [0.018 ; 0.024]	0.020 [0.016 ; 0.022]	0.030 [0.023 ; 0.031]	0.037 [0.029 ; 0.039]	0.041 [0.033 ; 0.043]
anticipé	0.077 [0.025 ; 0.124]	0.070 [0.021 ; 0.115]	0.102 [0.025 ; 0.168]	0.126 [0.035 ; 0.205]	0.141 [0.038 ; 0.230]
Permanent					
court terme	0.100 [0.033 ; 0.164]	0.092 [0.029 ; 0.150]	0.133 [0.035 ; 0.218]	0.165 [0.049 ; 0.273]	0.185 [0.050 ; 0.304]
long terme	0.164 [0.060 ; 0.291]	0.151 [0.052 ; 0.258]	0.218 [0.064 ; 0.384]	0.270 [0.091 ; 0.469]	0.303 [0.098 ; 0.533]

Notes : élasticités évaluées au ménage moyen de chaque zone, à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13. Intervalles de confiance à 95% calculés par bootstrap paramétrique.

A partir des paramètres du modèle rationnel estimés par MMG, la table 2.18 isole ces sensibilités en fonction de la zone de résidence : l'élasticité de l'automobilité des ménages à leur revenu est à court terme de +0.09 en Province jusque +0.18 à Paris, et de +0.15 jusque +0.30 à long terme. Lorsque le revenu courant augmente de façon exceptionnelle et non anticipée, l'élasticité de la mobilité varie de +0.06 à +0.12 de la Province jusque Paris, alors que la sensibilité estimée pour un accroissement de revenu exceptionnel anticipé varie de +0.07 à +0.14. Avec des mesures d'élasticité au revenu systématiquement comprises entre 0 et 1 quel que soit l'horizon, une croissance de revenu entraîne une augmentation proportionnellement moindre de l'automobilité des ménages, classant le bien kilomètre dans la catégorie des biens normaux nécessaires (VARIAN, 1995, p.120).

2.7.6 Effet de la localisation des ménages

Tous les modèles suggèrent la proximité, toutes choses égales par ailleurs, de l'automobilité des ménages de grande couronne francilienne et des ménages provinciaux : les tests d'égalité des paramètres du nombre d'automobiles dans ces zones sont systématiquement acceptés au seuil d'erreur à 5%. Même si on observe régulièrement dans les modèles de la table 2.13 que le kilométrage estimé d'une voiture parisienne est supérieur à celui estimé pour une voiture de petite couronne, tous réfutent l'hypothèse d'une différence significative entre ceux-ci. Néanmoins, les modèles estimés par MMG acceptent l'hypothèse d'une différence de kilométrage, *ceteris paribus*, entre une automobile de petite couronne et une autre de grande couronne. Ainsi, la frontière des comportements d'usage de l'automobile semble géographiquement se situer entre les deux couronnes de l'Île-de-France.

2.7.7 Effet du déménagement des ménages

La table 2.19 reporte les simulations symétriques de l'effet *ceteris paribus* d'un déménagement des ménages sur le kilométrage individuel de leurs voitures. A long terme, un ménage originaire de Province qui s'installe en grande couronne francilienne augmente en moyenne le kilométrage de chacune de ses voitures de 338 km, alors qu'il le diminue de 1877 ou de 3619 km selon qu'il va s'installer à Paris ou en petite couronne francilienne. Toutes autres choses égales, la relocalisation d'un ménage francilien de la grande vers la petite couronne

s'accompagne d'une diminution de 3957 km par voiture, et de 2215 km si sa destination est Paris. Enfin, un ménage parisien s'installant en petite couronne diminue le kilométrage de chacune de ses automobiles de 1742 km.

Même si elle n'est pas statistiquement significative, l'utilisation plus intensive de la voiture chez les Parisiens que chez les Franciliens de petite couronne pourrait s'expliquer par une distance plus importante les séparant d'activités situées en des zones qui rendent incontournable l'usage automobile, (par exemple, moins bien desservies en transports publics : en grande couronne ou en Province). Néanmoins, il faut remarquer que le déménagement d'un ménage le conduit significativement à reconsidérer son taux d'équipement : sur les mêmes données, nous montrerons dans le prochain chapitre l'effet très significatif du lieu de résidence d'un ménage sur son état de motorisation. Pour compléter cette analyse, le paragraphe 2.7.10 traite de l'effet sur l'automobilité des ménages induit par un changement de son niveau d'équipement.

TABLE 2.19 : Effets à long terme d'un changement de zone résidentielle :
Variation du kilométrage par automobile (en kilomètres)

Origine	Province	GC	PC	Paris
Destination				
Province	0	-338 [-3077; 1758]	3619 [1129 ; 6446]	1877 [-3885 ; 5624]
GC		0	3957 [1198 ; 8004]	2215 [-3619 ; 6391]
PC			0	-1742 [-8048 ; 1802]
Paris	<i>symétrique opposé</i>			0

Note : intervalles de confiance à 95% calculés par bootstrap paramétrique. Evaluations à partir du modèle d'addiction rationnelle estimé par MMG (table 2.12).

2.7.8 Effet du vieillissement du parc automobile

L'impact de l'âge des véhicules sur leur utilisation est très différent selon le carburant utilisé. Tous les modèles révèlent un effet très significatif et négatif

lorsque la voiture est à moteur essence, alors qu'il est statistiquement inexistant pour les voitures diesel⁵⁰. Théoriquement dans le modèle d'addiction rationnelle, l'effet du vieillissement des automobiles ne pourrait pas être considéré comme une « surprise » non anticipée permanente : puisque l'agent a pleinement conscience du vieillissement de son parc automobile avec les années, nous posons que l'effet d'une année additionnelle de l'âge d'une automobile est celui du long terme dès le court terme⁵¹. De plus, l'effet de court terme d'un changement permanent et renouvelé (comme ici les effets d'âge) à chaque période peut raisonnablement s'approximer par l'effet d'un changement permanent à long terme⁵². Le modèle estimé par MMG évalue à environ 300 km la réduction du kilométrage par voiture essence chaque année contre seulement une diminution annuelle de 35 km (non significative) pour les voitures à moteur diesel (table 2.20).

2.7.9 Effet des variables socio-démographiques du ménage

La position dans le cycle de vie du ménage est un déterminant explicatif de sa mobilité. Notamment, ceux dont le chef de famille est âgé de 18 à 39 ans « roulent » statistiquement plus que leurs aînés. En revanche, même si les ménages dont le chef est âgé de plus de 60 ans semblent avoir un usage moins important de la voiture que ceux dont le chef est âgé d'entre 40 et 60 ans, les modèles d'addiction rationnelle aboutissent à une différence non significative de l'automobilité entre ces deux catégories. A partir de ce modèle estimé par MMG, nous observons les effets du vieillissement du chef de famille. Le passage de la classe des « jeunes » à celle des « âges moyens » s'accompagne à long terme⁵³ d'une diminution significative du

⁵⁰ D'ailleurs réputées pour être bien plus robustes dans le temps que les voitures essence.

⁵¹ L'effet permanent de court terme est dans le modèle d'addiction rationnelle un effet de pure « surprise », alors que l'effet de long terme permanent calcule la variation totale induite par le changement permanent lorsque l'effet de surprise a été dissipé dans le temps.

⁵² Par exemple, considérons un cas discret où l'effet sur une variable Y d'un changement permanent de X est une variation de +1 pendant 3 ans, puis nulle ensuite. L'effet de long terme de ce changement permanent est donc +3. Si ce changement permanent est renouvelé de période en période, alors les effets cumulés sur une période donnent également +3 lorsque l'effet de long terme de la variation initiale est atteint (ici après 3 ans). Dans le cas continu, l'effet permanent de long terme peut ainsi approximer l'effet de court terme permanent cumulé.

⁵³ L'effet du changement de catégorie d'âge du chef de famille est difficilement interprétable à court terme puisque son vieillissement ne peut, là non plus, être pour lui une surprise non anticipée d'une part, et d'autre part puisque cet effet s'effectue vraisemblablement en continu avec les années (et non par paliers de tranches d'âges).

kilométrage d'environ 3650 km en moyenne, et de 650 km seulement (non significative) lorsque le passage s'effectue depuis cette classe d'âge intermédiaire vers celle des plus âgés.

TABLE 2.20 : Effets marginaux à court et long terme d'un changement permanent de déterminants sur l'automobilité des ménages (en kilomètres).

Horizon	Court terme	Long terme
Effet de l'ajout dans le ménage...		
...d'un permis de conduire	+2247 [645;3925]	+3687 [1109;7236]
...d'un adulte	+695 [-447;2074]	+1141 [-741; 4081]
...d'un actif occupé	+486 [-724;1468]	+797 [-1521;2214]
...d'une femme	-1965 [-3353;-571]	-3224 [-5879;-1070]
...d'un mineur	-504 [-1405;198]	-828 [-2763;341]
Effet du changement de tranche d'âge du chef de famille		
De (18-39 ans) vers (40-64 ans)	-2336 [-5036;-691]	-3832 [-10351;-1088]
De (40 -64 ans) vers (plus de 64 ans)	-392 [-1826;940]	-644 [-2995;1864]
Effet de l'augmentation ...		
... de 1000 Francs de revenu annuel	+12 [03;19]	+20 [06;34]
... de 1 Franc des prix de 100 kilomètres	-58 [-103;-12]	-95 [-182;-21]
Effet du vieillissement d'un an ...		
... d'un véhicule essence	-290 [-456;-122]	
... d'un véhicule diesel	-35 [-401;350]	

Notes : évaluations à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13. Intervalles de confiance à 95% calculés par bootstrap paramétrique.

L'effet d'un permis de conduire additionnel dans un ménage conduit globalement à une augmentation significative de son usage de la voiture (+2247 km à court terme, +3687 km à long terme), alors que la présence d'une femme adulte

plutôt qu'un homme tend à la diminuer significativement (-3224 km). Toutes autres choses égales, l'effet d'un adulte additionnel ou le remplacement d'un inactif par un actif occupé, n'augmente significativement pas l'automobilité du ménage, excepté dans le modèle d'addiction myope estimé par MMG. Enfin, les modèles révèlent unanimement que la présence d'enfants n'a pas d'influence statistique sur le kilométrage automobile des ménages. Ces résultats d'un changement permanent de caractéristiques du ménage sont résumés dans la table 2.20. *Ceteris paribus*, ces derniers ne sont pas surprenants : l'ajout d'un adulte dans un ménage correspond souvent à celui d'un enfant du chef de famille venant de passer l'âge de la majorité, et son effet reste non significatif comme l'est celui d'un mineur. L'effet positif sur l'automobilité d'un inactif devenu actif ne relève pas du simple changement de statut d'activité mais prendrait davantage sa source dans l'augmentation de revenu qu'il génère : le nouvel actif choisira par exemple de se rendre sur le lieu de son emploi (et ceux de ses loisirs) en voiture si notamment le surcroît de revenu apporté au ménage le lui permet, et bien entendu, s'il dispose du permis de conduire. Enfin, l'effet négatif sur l'automobilité des ménages d'une femme en remplacement d'un homme peut prendre son illustration dans des déplacements communs entre conjoints qui pourraient ne pas exister pas entre deux hommes adultes.

2.7.10 Effet de la motorisation / démotorisation des ménages

L'ajout d'une voiture dans le ménage peut être non anticipé si la décision de s'équiper davantage est prise seulement dans l'année en cours. Sous cette hypothèse, les effets théoriques de court terme dans le modèle d'addiction rationnelle prennent une signification. Les tables 2.22 et 2.23 reportent les effets estimés d'une automobile additionnelle⁵⁴ neuve ou d'occasion âgée de cinq ans. Ils tiennent compte des variations des sommes des caractéristiques (âges et coûts) qu'accompagne un agrandissement du parc du ménage (cf. notes des tables 2.22 et 2.23). Notamment, l'effet continu du vieillissement des automobiles oblige ici à fixer l'horizon temporel du long terme. Sans cela, l'effet en $t = +\infty$ d'une motorisation supplémentaire serait infiniment négatif compte tenu de l'accumulation des effets d'âge⁵⁵. Nous

⁵⁴ Le changement du nombre d'automobiles dans un ménage fait également varier dans notre spécification les variables d'âge total des véhicules ainsi que le prix des cent kilomètres. Ne pas en tenir compte lors du calcul des effets marginaux reviendrait à observer l'effet marginal d'une automobile neuve et gratuite à l'usage.

⁵⁵ Des automobiles à essence tout particulièrement (table 2.20).

choisissons de borner le long terme à 5 ans, dans la mesure où plus de 99% de l'effet total de long terme induit par un changement permanent de caractéristique est pris en compte par les ménages à cet horizon (table 2.21).

TABLE 2.21 : Pourcentage de l'effet de long terme pris en compte après t périodes

t	PC
0	61%
1	85%
2	94%
3	98%
4	99%
5	~100%

Lecture : lorsque le changement permanent est totalement non anticipé à la période courante, alors 61% de la variation de consommation induite à long terme est immédiatement (court terme) prise en compte par les ménages.

Note : calculs à partir de l'estimation MMG du modèle d'addiction rationnelle.

$$PC(t) = 1 - \varphi_2^{-(t+1)}.$$

En Province et grande couronne francilienne, l'ajout d'une automobile neuve à moteur essence vient augmenter d'environ 7300 km l'automobilité annuelle des ménages à court terme, et d'environ 10500 km à long terme. Pour une voiture diesel neuve supplémentaire, les augmentations respectives de court et de long terme sont d'environ 8400 km et 13600 km pour ces deux zones. Lorsque la voiture additionnelle est une occasion essence âgée de 5 ans, alors la mobilité des ménages augmente d'environ 6400 km à court terme et de 9100 km à long terme, et d'environ 8300 km et 13400 km aux horizons respectifs si le véhicule supplémentaire est diesel d'occasion (agé de 5 ans).

En petite couronne, l'effet estimé d'un véhicule neuf additionnel roulant à l'essence est une augmentation de l'automobilité du ménage de 4600 km à court terme et 6100 km à long terme, contre 6000 km et 9600 km pour une automobile diesel. Lorsque le véhicule ajouté est d'occasion, cette mobilité augmente à long

terme de 4700 km s'il consomme un supercarburant, et de 9400 km s'il fonctionne au gazole.

A Paris, une automobile supplémentaire neuve à moteur essence produit une augmentation de l'automobilité du ménage de 8600 km à long terme contre 11800 km pour une voiture additionnelle neuve de type diesel. Pour une occasion âgée de cinq ans, le surcroît de mobilité est estimé à 7100 km pour une voiture essence, et à 11600 km pour une voiture diesel.

TABLE 2.22 : Variations à court et long terme de l'automobilité annuelle du ménage par zone de résidence, induites par une automobile neuve additionnelle (en kilomètres)

Voiture neuve	Province	GC	PC	Paris
Court terme				
essence	+7247 [4014 ; 10219]	+7384 [4133 ; 10315]	+4604 [1117 ; 7721]	+6110 [1444 ; 10989]
diesel	+8344 [4775 ; 11366]	+8468 [4989 ; 11653]	+5967 [2401 ; 9167]	+7273 [3641 ; 11082]
Long terme à 5 ans				
essence	+10443 [7529 ; 13312]	+10668 [7049 ; 14986]	+6106 [1051 ; 10666]	+8578 [1346 ; 17902]
diesel	+13520 [10365 ; 16912]	+13722 [9978 ; 18367]	+9619 [4748 ; 14367]	+11762 [6842 ; 19050]

Notes : évaluations à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13. Les effets tiennent compte de l'augmentation inhérente de la variable $\sum_v PKM_v$, calculée à la moyenne des automobiles par type de carburant et par zone en 2000.

Ne pas en tenir compte reviendrait à simuler les effets d'une automobile supplémentaire pouvant rouler gratuitement.

TABLE 2.23 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par une automobile d'occasion additionnelle (en kilomètres)

Voiture d'occasion (5 ans à l'acquisition)	Province	GC	PC	Paris
Court terme				
essence	+6364 [3440 ; 8815]	+6501 [3510 ; 9254]	+3721 [523 ; 10676]	+5228 [766 ; 9944]
diesel	+8239 [4856 ; 11179]	+8363 [4942 ; 18392]	+5862 [2346 ; 9058]	+7168 [6825 ; 19092]
Long terme à 5 ans				
essence	+8994 [6238 ; 11618]	+9219 [5626 ; 13412]	+4657 [-351 ; 9145]	+7130 [9 ; 16377]
diesel	+13347 [9439 ; 17485]	+13549 [9002 ; 19104]	+9446 [3878 ; 14934]	+11590 [6404 ; 19199]

Notes : cf. notes du tableau 2.22.

Inversement, les tables 2.24 et 2.25 reportent les effets d'une démotorisation permanente et partielle⁵⁶ d'une automobile âgée de 5 et 10 ans sur le kilométrage annuel des ménages : ceux-là ne sont pas à court terme les opposés exacts de ceux d'une motorisation additionnelle puisque l'effet du vieillissement des automobiles qui venait diminuer progressivement le kilométrage des ménages avec le temps disparaît immédiatement avec la voiture qui sort du ménage. A long terme cependant, l'effet de la séparation d'une automobile âgée de 5 ans correspond à l'opposé de la variation d'automobilité induite à long terme (5 ans plus tard) par l'arrivée d'une automobile neuve dans un ménage.

En grande couronne francilienne et en Province, la disparition d'une automobile essence de 10 ans se traduit à long terme par une diminution d'automobilité d'environ 9100 km, et d'environ 13400 km pour une voiture diesel. En petite couronne francilienne, la diminution de long terme pour ces deux types de véhicule est respectivement de 4700 et 9400 km. A Paris, la diminution est plus importante : 7100 km pour une voiture essence et 11600 km pour une voiture diesel.

Notons que ces diminutions d'automobilité sont inférieures aux kilométrages annuels moyens des automobiles considérées (essence et diesel), suggérant que la démotorisation partielle s'accompagne d'un report d'automobilité sur les voitures restantes du ménage.

⁵⁶ Pour un ménage, cela signifie qu'il lui reste au minimum une voiture après qu'il se soit séparé définitivement d'une autre automobile.

TABLE 2.24 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par la démotorisation partielle d'une automobile de 5 ans (en kilomètres)

Démotorisation d'une voiture de 5 ans	Province	GC	PC	Paris
Court terme				
essence	-6364 [-8847 ; -3472]	-6501 [-9236 ; -3495]	-3721 [-6608 ; -515]	-5228 [-9952 ; -757]
diesel	-8239 [-11169 ; -4859]	-8363 [-11870 ; -4495]	-5862 [-9054 ; -2327]	-7168 [-10795 ; -3743]
Long terme				
essence	-10443 [-13334 ; -7404]	-10668 [-14901 ; -7027]	-6106 [-10626 ; -1042]	-8578 [-17880 ; -1360]
diesel	-13520 [-16857 ; -10365]	-13722 [-18773 ; -9165]	-9619 [-14359 ; -4746]	-11762 [-19033 ; -6840]

Notes : évaluations à partir de l'estimation MMG du modèle d'addiction rationnelle de la table 2.13. Les effets tiennent compte de la diminution inhérente de la variable $\sum_v PKM_v$, calculée à la moyenne des automobiles par type de carburant et par zone en 2000.

TABLE 2.25 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par la démotorisation partielle d'une automobile de 10 ans (en kilomètres)

Démotorisation d'une voiture de 10 ans	Province	GC	PC	Paris
Court terme				
essence	-5481 [-7662 ; -2924]	-5619 [-8186 ; -2864]	-2838 [-5595 ; 173]	-4345 [-8983 ; 10]
diesel	-8134 [-11271 ; -4557]	-8258 [-12307 ; -3556]	-5757 [-9247 ; -1938]	-7063 [-10792 ; -3533]
Long terme				
essence	-8994 [-11622 ; -6201]	-9219 [-13396 ; -5352]	-4657 [-9156 ; 329]	-7130 [-16339 ; 27]
diesel	-13347 [-17485 ; -9424]	-13550 [-19643 ; -7584]	-9446 [-14954 ; -3873]	-11590 [-19199 ; -6442]

Notes : cf. notes du tableau 2.24.

2.7.11 Effet des transitions 'essence \leftrightarrow diesel' lors du remplacement d'un véhicule

Notre modélisation permet également d'évaluer l'effet d'un remplacement d'une automobile essence par un véhicule diesel (et inversement) sur l'automobilité des ménages. Ce changement influe simultanément sur plusieurs facteurs explicatifs : les variables d'âge des véhicules essence et diesel, et la variable du prix du carburant pour cent kilomètres. Les caractéristiques des véhicules entrants et sortants lors du changement sont donc nécessaires aux calculs des effets. L'âge moyen des véhicules essence remplacés en 2000 par des véhicules diesel est d'environ 8.8 ans. Inversement, l'âge moyen du véhicule diesel remplacé en 2000 par une voiture essence est de 7.9 ans. Les consommations moyennes respectives reportées pour ces deux catégories de véhicules (aux âges arrondis) dans la vague Parc Auto 2000 sont respectivement de 7.7 et 6.9 litres de carburant aux cent kilomètres. Dans la même source, les âges moyens à l'acquisition des véhicules remplaçants sont de 3.1 ans pour le véhicule essence et 3 ans pour le véhicule diesel. La consommation kilométrique de ces nouvelles automobiles est respectivement de 7.4 et 6.9 litres de carburant consommés pour cent kilomètres.

TABLE 2.26 : Variation d'automobilité induit par le changement de motorisation d'un véhicule (en kilomètres)

Type de transition	Court terme	Long terme (à 5 ans)
essence vers diesel	+3558 [2520 ; 4727]	+4094 [2675 ; 5669]
diesel vers essence	-1605 [-2759 ; -506]	-3682 [-6784 ; -1994]

Note : calculs à partir de l'estimation MMG du modèle d'addiction rationnelle (table 2.13).

En multipliant ces variations de caractéristiques automobiles par leurs effets de court et long terme, on obtient les variations de kilométrage induit par un changement de type de motorisation. Toutes autres choses égales (notamment les prix des deux principaux carburants), l'effet moyen d'une transition « essence vers diesel » s'accompagne à court terme d'une augmentation moyenne de la mobilité

du ménage de 3558 km, et de 4094 km cinq ans plus tard. Inversement, la variation de kilométrage induit par le remplacement d'un véhicule diesel par un véhicule essence se traduit selon le modèle par une réduction de 1605 km à court terme et de 3682 km à long terme, 5 ans plus tard (table 2.26). Ces résultats confirment ceux obtenus par HIVERT (1999), et par HIVERT et CERRI (2004) à partir d'analyses descriptives sur des appariements tri-annuels d'observations, utilisant les vagues 1985 à 2001 de Parc Auto.

2.7.12 Limites et critiques de la démarche empirique

Bien entendu, l'usage des paramètres issus d'un estimateur reposant sur la variabilité inter-individuelle des observations (estimateur dit « *between* ») peut être critiquable lorsque l'on souhaite en dériver des élasticités de court et long terme, c'est-à-dire conclure sur une dynamique temporelle des comportements, et il aurait été sans doute préférable de calculer ces élasticités sur la base d'estimateurs longitudinaux (dits « *within* ») plutôt que transversaux. Mais la spécification du modèle d'addiction rationnelle est gourmande en vagues d'observations, et un estimateur longitudinal aurait impliqué de suivre les ménages durant au moins quatre années consécutives. Cela aurait été possible avec l'enquête Parc Auto mais au prix d'une attrition, qui, outre le fait d'être conséquente, aurait nécessité un nouveau calcul des pondérations individuelles. Cependant, un modèle correctement spécifié doit empiriquement aboutir à la proximité entre les résultats des estimateurs *within* et *between*⁵⁷. Sous cette hypothèse, utiliser les coefficients transversaux au service de conclusions dynamiques reste donc une démarche valide. Par ailleurs et compte tenu de la présence d'une seule variable de prix et de revenu dans les spécifications retenues pour l'ensemble de l'échantillon, la différenciation des élasticités par zone de résidence n'a pu être obtenue que sous le seul effet de moyennes spécifiques aux différentes régions. Nous avons bien tenté d'identifier des coefficients prix et revenu par zone de résidence, en élargissant le nombre de ces variables économiques par leur produit avec les indicatrices de localisation. Si les coefficients obtenus étaient proches⁵⁸ et de signe attendu pour la Province, la grande

⁵⁷ Voir MAIRESSE (1988) pour une illustration concernant les lois de production. Pour quelques postes de consommation, GARDES et al. (2005) montrent que les écarts observés résultent de biais d'endogénéité.

⁵⁸ Non statistiquement différents deux-à-deux.

et la petite couronne francilienne, la qualité des estimations allait toutefois en se dégradant dangereusement. Quant aux coefficients associés aux ménages parisiens, ils étaient même contre intuitifs, bien que non significatifs. Ces considérations (en partie dues à la diminution progressive des effectifs de ménages motorisés par zone depuis la Province jusque Paris, en passant par la grande, puis par la petite couronne francilienne) ont finalement justifié le calage des élasticités sur des coefficients « nationaux » de prix et de revenu.

Une autre critique que l'on pourrait faire à nos résultats repose sur le traitement de la variable de prix, construite sur la base d'un produit faisant notamment intervenir le type de motorisation (essence ou diesel) : celui du prix du carburant utilisé multiplié par la consommation déclarée des véhicules du ménage. En considérant cette variable de prix comme un input exogène à la détermination du kilométrage des ménages, nous avons implicitement émis l'hypothèse que les variables ayant servi à sa construction étaient également exogènes. Or, le choix du type de véhicule dans un ménage pourrait être déterminé en partie par son usage futur. Par exemple, les ménages roulant beaucoup peuvent préférer un équipement diesel pour profiter de la tarification avantageuse du gazole sur les supercarburants. De fait, notre variable de prix pourrait être endogène^{59,60}. Pourtant, la présence du prix kilométrique dans le lot des variables instrumentales n'avait pas conduit les tests de Hansen à rejeter l'hypothèse d'orthogonalité des instruments avec les aléas dans les modèles estimés par MMG, suggérant au contraire son exogénéité.

2.7.13 Comparaisons internationales

Les différentes considérations de prix entre études compliquent la démarche comparative entre pays : quelques auteurs considèrent le prix du kilomètre comme déterminant alors que d'autres utilisent directement le prix du carburant (GOODWIN, 2000). Dans ce chapitre, nous avons opté pour le premier choix, ceci

⁵⁹ Pour notre étude, ce n'est pas crucial. Puisque nous avons cherché à estimer l'effet « mécanique » total des prix, et non une mesure de microéconomie pure, nous n'avons pas à examiner la dynamique explicative sous-jacente de la demande de mobilité sur l'équipement automobile et le coût kilométrique. Il n'y a donc pas lieu d'instrumenter la variable des prix ; il est même important de la conserver telle.

⁶⁰ Notons que dans la littérature, les études cherchant à expliquer le kilométrage désagrégé à partir d'une moyenne pondérée des prix des deux principaux carburants peuvent être soumises à la même remarque.

pour des raisons de cohérence microéconomique, mais nos résultats entretiennent un rapport très étroit avec ceux utilisant le niveau de prix des carburants. En effet, sans évolution notable de la consommation kilométrique des automobiles, les effets de la variation du prix du kilomètre sur le trafic se résument aux effets du prix du carburant⁶¹. Puisque l'efficacité reste plutôt stable à court terme (table 2.7), la comparaison des élasticités pour ces deux types de prix peut être raisonnablement opérée à cet horizon. A long terme, la pratique est moins évidente. Comme nous sommes aujourd'hui dans une configuration générale de hausse des prix des carburants et d'une lente amélioration de l'efficacité des automobiles, l'évolution (en taux de croissance) de long terme du prix du kilomètre est sensiblement moins importante que celle du prix des carburants. Il en découle une élasticité du trafic automobile au prix du kilomètre plus élevée que celle au prix des carburants. La comparaison des effets revenu est également problématique : selon que la spécification retenue par les auteurs incorpore ou non une variable de stock automobile comme déterminant, les mesures de sensibilité du kilométrage sont très différentes puisque le revenu est un déterminant important du taux d'équipement (chapitre 3). La revue de littérature qui suit fait état d'une hétérogénéité des résultats relevant notamment des spécifications particulières utilisées dans chacune des études, compliquant davantage les opportunités de comparaisons.

HENSHER et al. (1990) se sont intéressés à la modélisation du kilométrage annuel des automobiles des ménages résidant dans l'aire urbaine de Sydney, à partir d'un échantillon d'environ 1200 observations couvrant la période 1981-1982. Leur spécification empirique entretient un lien de parenté assez étroit avec celle du présent travail : comme déterminants, les auteurs utilisent une série d'observations microéconomiques décrivant les attributs des automobiles possédées par les ménages, leurs caractéristiques sociales, économiques, démographiques et le coût calculé du carburant pour effectuer un kilomètre. Néanmoins, leurs données en coupe les contraignent à ne considérer qu'une spécification statique. Les résultats présentés en termes d'élasticités distinguent les effets-prix et du revenu selon le degré de motorisation des ménages. Dans le court terme⁶², leur fourchette de

⁶¹ Exemple : puisque le prix du kilomètre (en carburant) d'une automobile est défini comme le produit du prix du carburant utilisé avec sa consommation (en litres pour 100 kilomètres), une augmentation de 10% du prix du carburant se traduit également par une augmentation de 10% du prix du kilomètre si la consommation kilométrique est gardée constante.

⁶² Pour GRAHAM et GLAISTER (2002), les résultats reportés par HENSHER (1990) sont de court terme dans la mesure où les données utilisées ne couvrent qu'une année.

sensibilité aux prix du kilomètre varient de -0.22 pour les ménages mono-motorisés à -0.39 pour les ménages tri-motorisés dans les modèles incluant une correction du biais de sélectivité. Sans véritablement faire de références, HENSHER et al. notent que ces élasticités au prix sont très similaires à celles de travaux précédents. Pour notre cas d'étude français, on peut raisonnablement souligner la proximité de leurs élasticités de court terme avec celles obtenues pour la région parisienne dans la table 2.15 lorsque l'on sait que la majorité des ménages franciliens motorisés ne possèdent qu'une seule automobile (72% d'entre eux en 2000). Par ailleurs, les auteurs remarquent la proximité des sensibilités du kilométrage des voitures aux variations du coût kilométrique en carburant et du coût par kilomètre de l'entretien courant des véhicules. Ce résultat intéressant tend à suggérer que les agents prennent en considération un coût « global » au kilomètre pour déterminer leur usage automobile.

Les auteurs notent cependant une faible sensibilité (voire statistiquement inexistante) du kilométrage au revenu net des frais automobiles : selon le degré de motorisation, les élasticités au revenu s'échelonnent de +0.05 à +0.14. Cités à titre de comparaison dans leur papier, GREENE et HU (1984), MANNERING et WINSTON (1985) ont respectivement estimé des élasticités du kilométrage au revenu de +0.13 et +0.11 pour les Etats-Unis. Avec des valeurs estimées comprises entre +0.08 et +0.15 selon les modèles utilisés (table 2.17), nos résultats sont également très proches de ceux obtenus par ces derniers auteurs.

OUM et al. (1992) reportent des élasticités prix directes pour la « demande d'usage automobile » issues de sept travaux empiriques variés de la seconde moitié des années 1980. Pour cette période et aux Etats-Unis, l'élasticité prix de court terme est estimée à -0.23, et à -0.28 à long terme. Aux mêmes horizons, les élasticités australiennes reportées varient respectivement de -0.09 à -0.24, et de -0.22 à -0.31. Au Royaume-Uni, la sensibilité estimée renseignée par les auteurs est comprise entre -0.14 et -0.36 sans horizon temporel précis. Si ces résultats s'approchent de ceux de la table 2.14 pour le court terme, nos élasticités de long terme semblent toutefois légèrement supérieures. Les auteurs font néanmoins constat d'élasticités de long terme peu significativement supérieures à celles de court terme. Pour l'expliquer, les auteurs suggèrent que les études répertoriées ne « développent pas toutes de véritables modèles de long terme qui prennent en compte les changements de motorisation et les choix de localisation ».

ELTONY (1993) se démarque des études précédentes en reliant la consommation de carburant par automobile au prix du carburant par mile et au revenu des ménages, sur des données semi-agrégées des régions canadiennes entre 1969 et 1988. Sous l'hypothèse que la consommation kilométrique en carburant des automobiles est stable à court terme, alors les élasticités prix à cet horizon correspondent à celles que nous avons considérées dans notre étude (les quantités de carburant étant *a priori* proportionnelles au kilométrage effectué). La spécification d'ELTONY est de type log-log statique pour son équation de consommation de gallons de carburant par les ménages, et il en découle des élasticités prix et revenu du ménage respectives de -0.21 et +0.15 ; des valeurs tout à fait comparables aux nôtres. Ce n'est là qu'un résultat partiel de l'auteur qui développe en outre tout un modèle qui mesure notamment l'effet de la hausse du prix des carburants sur la consommation énergétique des véhicules et le volume du parc automobile en circulation. Il en résulte que la sensibilité totale à court terme lorsque l'on considère également ces effets est une élasticité prix de -0.31. Ainsi, ELTONY met en évidence qu'une forme structurelle selon qu'elle tient compte ou non du stock d'automobiles notamment peut faire varier les résultats des élasticités-prix.

GRAHAM et GLAISTER (2002) synthétisent les résultats d'études traitant des élasticités de demande automobile aux tarifs des carburants. Les recherches de GOODWIN (1992) dans ce domaine y sont référencées : sur la base de quatre élasticités obtenues dans des travaux conduits dans les années 1980, l'auteur reporte une sensibilité moyenne du trafic automobile au prix des carburants de -0.16 à court terme et de -0.33 à long terme sur séries temporelles. Les écarts-types associés à ces moyennes révèlent cependant une hétérogénéité des résultats obtenus : respectivement 0.08 et 0.11 pour ces deux horizons. Sur deux autres études estimant cet effet en coupe instantanée, la moyenne des élasticités répertoriée est de -0.29.

GOODWIN et al. (2004) mettent à jour ces résultats sur la base de travaux empiriques menés depuis les années 1990 sur les données de pays aux caractéristiques comparables à celles du Royaume-Uni. Dans leur document, les auteurs séparent les élasticités prix et revenu issues des modèles dynamiques et statiques. Pour la première classe de modèles, la sensibilité moyenne du kilométrage au prix du carburant calculée par les auteurs est de -0.10 à court terme, et de -0.30 à long terme. Mais comme précédemment, la variabilité autour de ces moyennes (mesurée par un écart-type) est assez élevée : 0.06 et 0.25. Si la valeur de long

terme reste comparable avec celles répertoriées dans la table 2.14, celle de court terme paraît nettement moins élevée que les nôtres.

JOHANSEN et SCHIPPER (1997) ont utilisé des données macroéconomique de 12 pays de l'OCDE (parmi lesquels figure la France) entre 1973 et 1992 pour notamment estimer une équation dynamique de la distance moyenne parcourue par les automobiles en fonction du coût du kilomètre en carburant, du revenu, du taux d'automobile *per capita*. On retient notamment de leur document que les sensibilités mesurées varient selon la méthode d'estimation mise en oeuvre. Sur la base des paramètres les plus « raisonnables », les auteurs reportent des élasticités prix et revenu à long terme respectives de -0.2 et +0.2 pour l'ensemble de leur échantillon.

FRIDTSRÖM (1998) utilise des données semi-agrégées mensuelles des contés norvégiens entre 1973 et 1994 pour expliquer l'usage automobile notamment en fonction du revenu et du coût d'utilisation avec un modèle à ajustement partiel. En tenant compte des effets sur le stock d'automobile détenu, les élasticités au revenu et au prix des carburants que l'auteur reporte sont respectivement de +0.38 et -0.24.

BERRI et al. (2005) ont évalué l'effet des variables économiques sur l'usage automobile des ménages moyens de cohortes, sur des données françaises couvrant la période 1977-1994. Les auteurs estiment notamment à +0.55 l'élasticité de l'automobilité à la consommation finale des ménages (en proxi du revenu). Parce que leur résultat ne maintient pas constantes des variables positivement corrélées avec l'automobilité, telles que les nombres d'automobiles et de permis de conduire, ce résultat apparaît très logiquement supérieur à notre estimation (+0.10 et +0.16 à court et long terme, table 2.17). Par ailleurs, l'effet reporté par les auteurs du prix des carburants sur l'automobilité est une élasticité de -0.09, une valeur faible au regard de nos résultats présentés dans la table 2.14. Même si les prix considérés dans les deux études n'obéissent pas à la même définition, l'écart de sensibilité observé tient notamment à la nature différente des données manipulées. En effet, nous considérons dans notre étude des observations désagrégées de ménages motorisés, alors que les auteurs utilisent des ménages moyens par cohorte qui incorporent des ménages non-motorisés, fatalement plus insensibles au prix des carburants. C'est donc logiquement que nous obtenons une élasticité-prix de l'automobilité des ménages négativement plus élevée dans notre travail. Plus simplement, il se pourrait également que les sensibilités des ménages aux prix soient

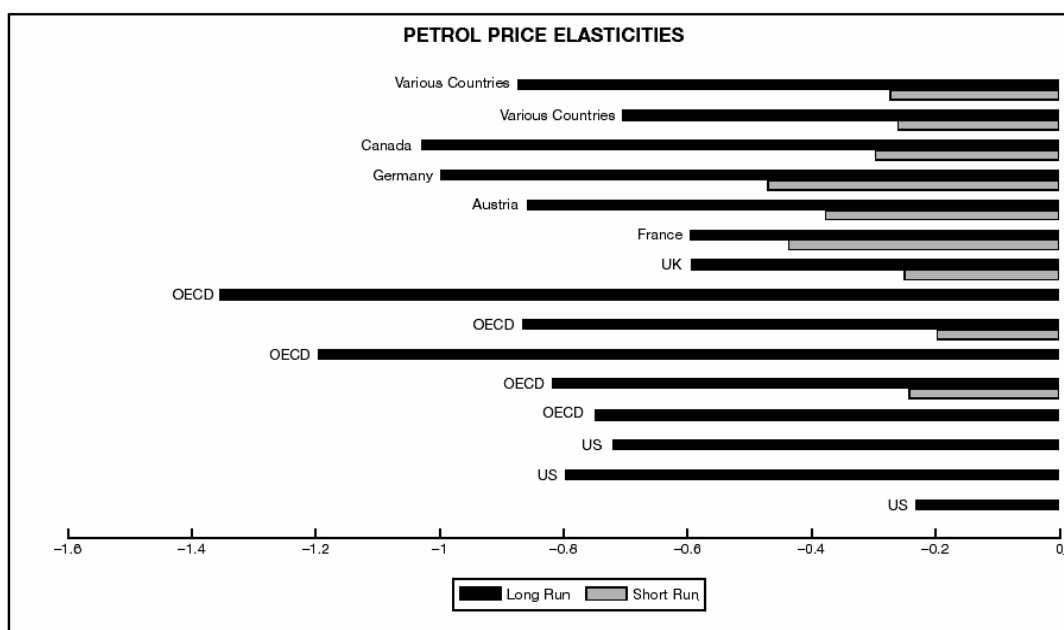
plus élevées dans le contexte de notre période d'observation que dans celui de BERRI et al. (2005).

Finalement, même si toutes les études concluent à une inélasticité⁶³ de la demande de carburant ou de l'usage automobile par rapport aux différents prix considérés, l'amplitude des écarts entre les résultats des travaux peut quelques fois surprendre. Naturellement, ce constat peut s'expliquer en partie par des analyses d'échantillons couvrant différents pays à différentes époques, mais aussi par des spécifications de modèles et des méthodes d'estimation différentes.

Pour notre cas particulier, le modèle le plus enclin à comparaisons avec les autres sources de résultats est certainement le modèle d'addiction myope puisque la majorité des équations dynamiques qui précèdent le report d'élasticités de court et de long terme dans la littérature repose sur une spécification autorégressive d'ordre un ($AR(1)$). A ce titre, l'élasticité-prix de long terme du modèle d'addiction myope, estimée à -0.3, semble être une valeur très raisonnable (voire consensuelle). Relativement, l'élasticité obtenue pour ce modèle à court terme (-0.22) peut apparaître élevée, témoignant d'un processus d'adaptation aux variations de prix particulièrement rapide en France. Dans la figure 2.3 qui suit, GRAHAM et GLAISTER (1992) reportent les élasticités prix de la demande de carburant pour une série de pays occidentaux : en comparaison, la France possède l'une des plus forte sensibilité à court terme et l'une des moins élevées à long terme, avec un ratio d'environ 0.75, alors qu'il est sensiblement inférieur à 0.5 dans les autres pays européens examinés (Allemagne, Royaume-Uni, Autriche). Par ailleurs, HIVERT (2004) a fait l'analyse de l'évolution du kilométrage entre 1999 et 2001 pour des ménages français ne modifiant ni leur localisation résidentielle, ni leur revenu, ni leur taux de motorisation. En contrôlant ces variables, l'objectif de la démarche était d'obtenir l'effet « propre » de l'évolution du prix des carburants sur l'automobilité des ménages avec un raisonnement de type *ceteris paribus*. L'auteur obtient une élasticité de court terme de -0.23, similaire aux résultats des modèles d'addiction myope et rationnels estimés par MMG de la table 2.14.

⁶³ L'« inélasticité » désigne des valeurs d'élasticité comprises entre 0 et 1 en valeur absolue.

FIGURE 2.3 : Mesures internationales d'élasticités-prix de la demande de carburant



Source : GRAHAM et GLAISTER (1992).

2.8 Conclusion

Après avoir montré que le kilomètre pouvait être traité comme un bien économique traditionnel, nous avons mis en application différentes équations de demande. Plus qu'un exercice de modélisation pour expliquer les kilométrages individuels des ménages, ce chapitre voulait également apporter quelques éclairages à la thématique de la dépendance automobile. C'est pourquoi les résultats présentés mettent l'accent sur le modèle d'addiction rationnelle de BECKER, GROSSMAN et MURPHY (1994), qui n'avait pas encore été testé sur des données de transport. Celui-ci apparaît plutôt pertinent pour décrire le comportement empirique d'automobilité des ménages français en 2000. En effet, l'addiction à l'usage de l'automobile est statistiquement avérée, si bien qu'on ne peut réfuter les propos de DUPUY (1999) et WICKHAM (2002) cités en introduction. Ensuite, nos résultats accréditent la version rationnelle de l'addiction lorsque le modèle est estimé par la méthode des moments généralisés (MMG) et le taux de substitution intertemporel

qui en découle est de 17%, une valeur plausible⁶⁴. Les agents seraient donc théoriquement conscients de la dynamique temporelle de leur consommation de kilomètres en voiture, et tiendraient compte des effets futurs de leur mobilité actuelle pour déterminer leur kilométrage annuel optimal. Cette conclusion originale démarque notre travail des précédentes études dont les spécifications économétriques introduisaient la dynamique des comportements uniquement grâce à une variable endogène retardée. Celles-ci sont très pratiques pour déduire des élasticités de court et long terme, mais conditionnent implicitement les agents à réagir de façon myope au sens de BECKER et al. (1994) c'est-à-dire sans considération de leur situation future. Néanmoins, la spécification $AR(1)$ du modèle d'addiction myope est également estimée et reportée à des fins de comparaison avec les autres études.

A l'échelle nationale, les élasticités de l'automobilité au prix kilométrique et au revenu, calculées sur les estimations MMG du modèle d'addiction rationnelle, ne font pas apparaître de réelle surprise, malgré sa spécification inhabituelle. Ces mesures respectives sont de -0.23 et +0.10 à court terme, -0.37 et +0.16 à long terme, et sont d'amplitude comparable avec celles d'autres travaux empiriques qui utilisent une spécification voisine de la nôtre. En distinguant le type de motorisation, nous remarquons que les voitures essence sont plus sensibles que les voitures diesel aux variations de prix de leurs carburants. Nous calculons que l'élasticité à court terme du kilométrage d'une voiture diesel au prix du gazole est de -0.13, et que celui d'une voiture essence au prix des supercarburants est de -0.32. Les sensibilités de long terme sont environ 2.5 fois plus fortes.

A l'échelle régionale, nous nous sommes attachés dans ce chapitre à mettre en perspective quelques résultats spécifiques à la région Île-de-France. Notre effort s'est notamment porté sur la distinction des effets du revenu et des prix des deux carburants principaux (supers et gazole) sur le kilométrage annuel des ménages franciliens. Depuis la Province, en passant par la grande puis, la petite couronne francilienne, jusque Paris, les ménages rendent compte d'une sensibilité au revenu croissante de façon quasi linéaire. La sensibilité au prix kilométrique augmente progressivement jusqu'en petite couronne francilienne, où celle-ci devient comparable avec celle des ménages parisiens.

⁶⁴ Et proche de celle obtenue par GARDES et al. (2007) dans leur analyse des dépenses de transport en Pologne.

Séparer les élasticités selon le type de carburant donne lieu aux observations suivantes. Les voitures essence les plus sensibles aux prix des supercarburants se trouvent à Paris et en petite couronne francilienne. Cela semble assez naturel puisque les résidents de ces zones bénéficient d'offres alternatives performantes par les transports publics. Pourtant, nous ne retrouvons pas la même configuration concernant les véhicules diesel. A Paris, ces automobiles présentent même la plus faible des sensibilités aux prix de leur carburant (-0.1 à court terme, -0.2 à long terme). Pour l'expliquer, nous pouvons avancer l'argument suivant. Ces ménages parisiens ont une forte mobilité, dont on imagine qu'elle s'oriente vers des régions éloignées en Province et peu desservies par les transports publics, ce qui justifie leur choix d'un véhicule diesel pour bénéficier d'un carburant moins cher à la pompe.

Chapitre 3 :

*Dynamique du niveau de motorisation
des ménages*

3.1 Introduction

Le nombre d'automobiles dans un ménage a un impact incontestable sur sa mobilité, tant sur un plan qualitatif, comme le choix des modes de déplacement (UNCLES, 1987), que sur un plan quantitatif, comme le nombre et les distances de déplacements (MEURS, 1990). C'est pourquoi la modélisation du comportement d'équipement en automobiles des ménages devient une nécessité lorsque l'on veut comprendre et prévoir leur demande de déplacement. Il s'agit là d'une préoccupation de premier rang pour la gestion de l'environnement et des infrastructures routières, pour l'industrie automobile et pour les services d'assurances... La place de l'automobile dans les pays développés et ses problématiques d'intérêt croissant ont indirectement accru les besoins de modélisation et ont accéléré la constitution de sources de données. Quelques unes sont structurées en panel, c'est-à-dire qu'elles permettent le suivi d'un échantillon de ménages au cours du temps : le « Dutch National Mobility Panel » aux Pays-Bas, le « Sydney Automobile Panel » en Australie, le « Mobilität Panel » en Allemagne, ou encore « Parc Automobile » en France en sont quelques exemples pour le transport. NOBILE et al. (1997), ou encore KITAMURA et BUNCH (1990) ont ainsi modélisé la possession automobile sur ce type de données. Cette double dimension des panels est un atout d'analyse certain par rapport aux enquêtes « unidimensionnelles » (séries temporelles ou coupes transversales) puisqu'elle permet de considérer conjointement la dynamique et l'hétérogénéité observable et non observable des comportements (SEVESTRE, 2003, p. 3). Cette structure de données semble donc avantageuse pour modéliser le taux de motorisation individuel puisqu'il nous vient plusieurs motifs élémentaires faisant dépendre le niveau courant d'équipement automobile du niveau de motorisation passé. Par exemple, l'automobile est à l'échelle du ménage un bien durable dont l'équipement n'est pas fréquemment remis en question. C'est aussi un bien relativement peu flexible, dont les quantités observées peuvent s'ajuster avec retard aux quantités désirées, notamment à cause du temps de recherche pour l'acquisition ou la cession (GOODWIN et MOGRIDGE, 1981). Egalement, l'automobile est potentiellement un bien dont on prend l'habitude de l'usage (chapitre 2), ce qui tend à maintenir le taux d'équipement individuel... Du point de vue de la modélisation statistique, cette potentielle dépendance temporelle des comportements de motorisation requiert l'usage d'une spécification économétrique dynamique, qui intègre donc des variables

retardées. Parmi celles-ci, la plus naturelle est une variable de motorisation passée qui devient endogène s'il existe par ailleurs un facteur d'hétérogénéité inobservable. Dans les modèles statistiques de réponses qualitatives ordonnées à base de variables latentes, l'existence de ces deux déterminants réclame des outils d'estimation relativement complexes, comme le maximum de vraisemblance approximée utilisant des points de quadrature (BUTLER et MOFFIT, 1982), ou le maximum de vraisemblance simulée (GOURIEROUX et MONFORT, 1994). Dans la littérature des transports, quelques exemples de modélisation dynamique portant sur l'état de motorisation d'observations désagrégées et utilisant ces techniques sont KITAMURA et BUNCH (1990), HANLY et DARGAY (2000), DARGAY et al. (2006). Une solution pratique pour conserver les méthodes d'estimation simples consiste à ne considérer qu'un seul des deux effets, en posant l'hypothèse que l'autre n'existe simplement pas. Par exemple, HANLY et DARGAY (2000) estiment également des modèles sans effet d'hétérogénéité inobservable. Mais idéalement, la modélisation doit tenir compte conjointement de ces deux effets.

Dans ce chapitre, nous utilisons un modèle probit ordonné dynamique avec hétérogénéité pour décrire le degré de motorisation automobile individuel en France. Trois catégories de ménages sont considérées : les ménages non-motorisés, ceux mono-motorisés, et les multi-motorisés. Le modèle est ajusté sur les données du panel pondéré 1999-2001 de l'enquête Parc Auto, dont la constitution a été décrite dans le premier chapitre.

Alors que les études précédentes utilisent un ensemble d'indicateurs du niveau (observé) de motorisation passée comme facteurs explicatifs dynamiques, notre modélisation probit s'en démarque et s'appuie sur une spécification latente linéaire autorégressive, inspirée de PAAP et FRANSES (2000). Cette formulation particulière du modèle latent nous permet d'étendre les propriétés d'analyse des modèles linéaires dynamiques aux outils d'interprétation du probit ordonné, sous la forme d'effets probabilistes marginaux de court et long terme. Par ailleurs, une forme flexible réduite, empruntée à l'approche d'HECKMAN (1981) et à l'estimateur linéaire dynamique de BLUNDELL et SMITH (1991), est utilisée pour approximer les conditions initiales du modèle latent.

Dans les deux premières sections, nous justifions et exposons le modèle statistique utilisé pour notre étude, puis dérivons ses outils d'interprétation. Dans une troisième partie, nous situons le modèle dans un cadre économétrique bayésien, et l'estimons à l'aide de l'échantillonneur de Gibbs (augmenté), une méthode de

Monte Carlo par chaînes de Markov particulière. Nous observerons ensuite quelques statistiques descriptives du panel Parc Auto, relatives à la motorisation des ménages dans une quatrième section, avant de commenter dans une cinquième partie l'ensemble des résultats d'estimation. La dernière section est consacrée à leur exploitation. Nous mesurons la vitesse d'ajustement du comportement des ménages aux variations de leurs caractéristiques, et nous examinons les effets de court et long terme d'un déménagement dans une zone francilienne (Paris, petite ou grande couronne) sur les probabilités de motorisation de ménages synthétiques. Enfin, les élasticités au revenu des niveaux d'équipement automobile à court et long terme sont présentées.

3.2 Le modèle catégoriel probit ordonné dynamique

Dans cette première section, nous décrivons le calque méthodologique que nous avons déployé sur les ménages de notre panel pour analyser leur choix de niveau d'équipement automobile. Bien que d'autres calques puissent évidemment exister, nous choisissons de conduire cette étude avec l'outil statistique.

Il convient en premier lieu de choisir dans la panoplie des modèles économétriques celui qui semble le mieux adapté pour décrire une variable quantitative d'équipement. Usuellement, ce choix est dicté par la nature de la variable à expliquer. Dans l'enquête Parc Auto, nous disposons individuellement pour chaque ménage du nombre d'automobiles à sa disposition, et pour caractériser cette variable quantitative discrète, à valeurs positives, un modèle de comptage de type Poisson semble théoriquement le plus adapté. Mais au regard de la distribution des ménages par nombre de voitures, il apparaît que 95% des ménages entre 1999 et 2001 possèdent aucune, une ou deux automobiles dans leur parc. Autrement dit, les effectifs de ménage possédant trois automobiles et davantage sont relativement réduits : 4.7% en 1999, 4.6% en 2000, 5.1% en 2001 (HIVERT, 2001 ; SOFRES 2001, 2002), et l'ajustement d'un modèle de Poisson, théoriquement défini sur $\mathbb{N} = \{0, 1, \dots, +\infty\}$, sur des observations quasi totalement distribuées sur les trois premières modalités de comptage $\{0, 1, 2\}$ peut au final sembler faiblement approprié. C'est pourquoi un modèle économétrique dont le nombre possible d'output est ajustable et limité paraît plus adapté à nos observations : cela nous amène à rassembler en tranches des modalités de comptage mitoyennes avec des effectifs trop faibles.

Pratiquement, nous avons conservé la classe des ménages non-équipés, celle des ménages équipés d'une seule voiture, et nous avons agrégé les ménages disposant de deux voitures ou plus en une classe de ménages multi-motorisés. La variable que nous souhaitons modéliser est donc ici catégorielle avec trois modalités, les ménages ayant été ventilés en trois ensembles mutuellement exclusifs : non-motorisés, mono-motorisés, et multi-motorisés. Compte tenu de l'ordonnancement logique de ces trois niveaux de motorisation, selon la quantité d'automobiles à disposition, un calque statistique adapté pour la représentation de cette variable (discrète et ordonnée) peut être choisi dans la famille des modèles de choix discrets ordonnés.

3.2.1 Présentation du modèle

Une bonne façon d'introduire les modèles catégoriels ordonnés est de partir des observations. Notons Y_{it} la variable du niveau de motorisation pour le ménage i à la période t , qui peut dans notre cas prendre trois valeurs ($Y_{it} = j \mid j \in \{0, 1, 2\}$), puisque sont pris en compte trois états de motorisation :

$$\begin{aligned} Y_{it} &= 0 \text{ si le ménage } i \text{ est non-motorisé à la période } t, \\ Y_{it} &= 1 \text{ s'il est mono-motorisé,} \\ Y_{it} &= 2 \text{ s'il est multi-motorisé.} \end{aligned} \tag{3.1}$$

Y_{it} est donc une variable codée issue de la motorisation observée du ménage, que les modèles de choix ordonnés supposent généralement résulter de la discrétisation d'une variable latente Y_{it}^* , continue et inobservée, obéissante à la règle suivante :

$$\begin{aligned} Y_{it} &= 0 \text{ si } -\infty < Y_{it}^* \leq s_1, \\ Y_{it} &= 1 \text{ si } s_1 < Y_{it}^* \leq s_2, \\ Y_{it} &= 2 \text{ si } s_2 < Y_{it}^* \leq +\infty, \end{aligned} \tag{3.2}$$

avec $\{s_1, s_2\}$, le couple des seuils inconnus qui règlent ainsi la classification des ménages dans l'un des trois états de motorisation. Le système (3.2) est la clé de passage entre variables latentes inobservées Y^* et variables « d'output » observées,

et pose la base d'un modèle de mesure (XIANG, 1989) dont la spécification économétrique porte uniquement sur Y^* . Les données de panel permettent l'élargissement de spécifications, usuellement statiques, à des configurations dynamiques souvent plus réalistes et dont les propriétés élargissent de fait l'éventail des conclusions. Pratiquement, une spécification est généralement rendue dynamique lorsque sont introduites dans sa formulation des variables endogènes et explicatives retardées. Au premier ordre, le modèle latent peut s'écrire sous la forme d'un modèle autorégressif à retards distribués, noté $ARDL(1,1)$:

$$Y_{it}^* = X_{it}\beta + \gamma Y_{it-1}^* + X_{it-1}\psi + \xi_{it}, \quad (3.3)$$

où X_{is} est la matrice traditionnelle des facteurs explicatifs du ménage i à la période s , Y_{it-1}^* est sa variable endogène retardée, $\{\beta, \gamma, \psi\}$ est un ensemble de paramètres, et ξ_{it} est le terme d'erreur aléatoire. Techniquement, la stationnarité du processus (3.3) nécessite notamment que le coefficient γ réside dans le cercle unitaire. Economiquement, l'interprétation des comportements est impossible pour $\gamma < 0$. Ces deux conditions réunies imposent finalement que $0 \leq \gamma < 1$. La spécification (3.3) peut être réécrite sous la forme d'un modèle vectoriel à correction d'erreur, noté VECM⁶⁵. En lui retranchant Y_{it-1}^* de chaque côté, puis en lui ajoutant et retirant $X_{it-1}\beta$ à droite de l'égalité, nous obtenons effectivement :

$$\Delta Y_{it}^* = \Delta X_{it}\beta + \lambda[Y_{it-1}^* - X_{it-1}\varphi] + \xi_{it}, \quad (3.4)$$

avec $\lambda = (\gamma - 1)$, et $\varphi = -(\psi + \beta)/\lambda$. Cette formulation VECM en (3.4) se compose d'une relation d'équilibre, $\Delta Y_{it}^* = \Delta X_{it}\beta$, et d'une erreur de déséquilibre à la période passée : $er_{it-1} = Y_{it-1}^* - X_{it-1}\varphi$. Lorsque cette dernière est positive, elle a un effet compensatoire négatif sur Y_{it}^* puisque $\lambda < 0$. Au contraire, elle induit un effet de rattrapage positif sur Y_{it}^* lorsqu'elle est négative. Le paramètre λ agit donc comme un ressort qui permet une correction d'autant plus rapide (ou lente) que γ tend vers 0 (ou 1) : il reflète donc la vitesse d'ajustement de Y^* à X dans leur relation à long terme.

Ce format d'écriture est particulièrement avantageux puisque les paramètres β et φ mesurent respectivement les effets de court et long terme d'un changement permanent de X_i sur Y_i^* . Par mesure de simplicité, nous posons dans cette étude la

⁶⁵ Acronyme anglosaxon pour « vectorial error correction model ».

restriction $\psi = 0$ et faisons ainsi l'hypothèse que le modèle (3.3) n'est dynamique que par la présence de Y_{it-1}^* . Littéralement, ceci revient à dresser une spécification pour Y_{it}^* qui est une relation linéaire des variables caractéristiques contemporaines, de l'endogène retardée, et du terme d'erreur :

$$Y_{it}^* = X_{it}\beta + \gamma Y_{it-1}^* + \xi_{it}, \quad (3.5)$$

et qui pose une relation de proportionnalité entre effets de court et long terme β et φ , dont le facteur dépend de γ : $\varphi = (1 - \gamma)^{-1} \beta$.

3.2.2 Décomposition de l'erreur et paramétrage

Les données de panel nous permettent la prise en considération d'un effet d'hétérogénéité inobservable, noté α_i , individuel et invariant dans le temps. Ainsi, nous pouvons détailler l'erreur globale ξ_{it} en deux composantes :

$$\xi_{it} = \alpha_i + \varepsilon_{it}, \quad (3.6)$$

où les termes d'hétérogénéité α_i sont traités comme des effets aléatoires, tandis que ε_{it} est le terme d'erreur idiosyncratique. Les hypothèses relatives à leur distribution sur la population et à leurs propriétés d'indépendance sont supposées les suivantes :

$$\begin{aligned} \alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2) & \alpha_i \perp \alpha_j \quad \forall i \neq j, \\ \varepsilon_{it} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) & \varepsilon_{it} \perp \varepsilon_{js} \quad \forall [i, t] \neq [j, s]. \end{aligned} \quad (3.7)$$

Puisque α_i et ε_{it} sont des aléas supposés normaux et indépendants, alors le terme ξ_{it} est normalement distribué :

$$\begin{aligned} \xi_{it} &\sim \mathcal{N}(0, \sigma_\xi^2), \\ \sigma_\xi^2 &= \sigma_\alpha^2 + \sigma_\varepsilon^2. \end{aligned} \quad (3.8)$$

Ce paramétrage particulier des erreurs range notre modèle de choix ordonnés dans la famille des modèles « probit ». En comparaison, l'usage d'une loi logistique pour ξ_{it} l'aurait placé dans la gamme des modèles « logit ». Ici, l'utilisation d'un probit tient plutôt à des conventions dans notre domaine d'étude qu'à des

préoccupations techniques. Pour les choix ordonnés effectivement, les modèles probit sont davantage utilisés en sciences sociales que leurs concurrents logit, mais on note que c'est l'inverse en biostatistiques par exemple (POWERS et XIE, 2000, p. 215).

3.2.3 Identification du modèle

Sauf à poser quelques restrictions courantes, les modèles probit ne sont pas identifiables. Pour s'en rendre compte, nous pouvons observer l'illustration graphique de la figure 3.1, avec $\phi(Y_{it}^* | X_{it}, Y_{it-1}^*)$ la fonction de densité conditionnelle de Y_{it}^* qui est celle d'une $\mathcal{N}(X_{it}\beta + \gamma Y_{it-1}^*, \sigma_\xi^2)$ compte tenu de la spécification (3.5) et le paramétrage (3.8).

Par exemple, intéressons nous à l'état de mono-motorisation de l'agent i à la période t dont la probabilité est dans le panneau (a) de la figure 3.1 mesurée par la surface grisée entre s_1 et s_2 , calculée par $\Pr(Y_{it} = 1) = \int_{s_1}^{s_2} \phi(Y_{it}^* | X_{it}, Y_{it-1}^*) dY_{it}^*$. Une translation horizontale linéaire de $\pm a$ sur l'axe des abscisses de la fonction de densité et de ses seuils ne modifie pas sa probabilité d'être mono-équipé : $\Pr(Y_{it} = 1) = \int_{s_1 \pm a}^{s_2 \pm a} \phi(Y_{it}^* \pm a | X_{it}, Y_{it-1}^*) d(Y_{it}^* \pm a)$ (panel (b) pour $c = 1$). Une première condition d'identification, consiste à ancrer un seuil sur une valeur fixe pour s'affranchir du problème de translation. Ainsi, nous posons par convention :

$$s_1 = s_1 \pm a = 0. \quad (3.9)$$

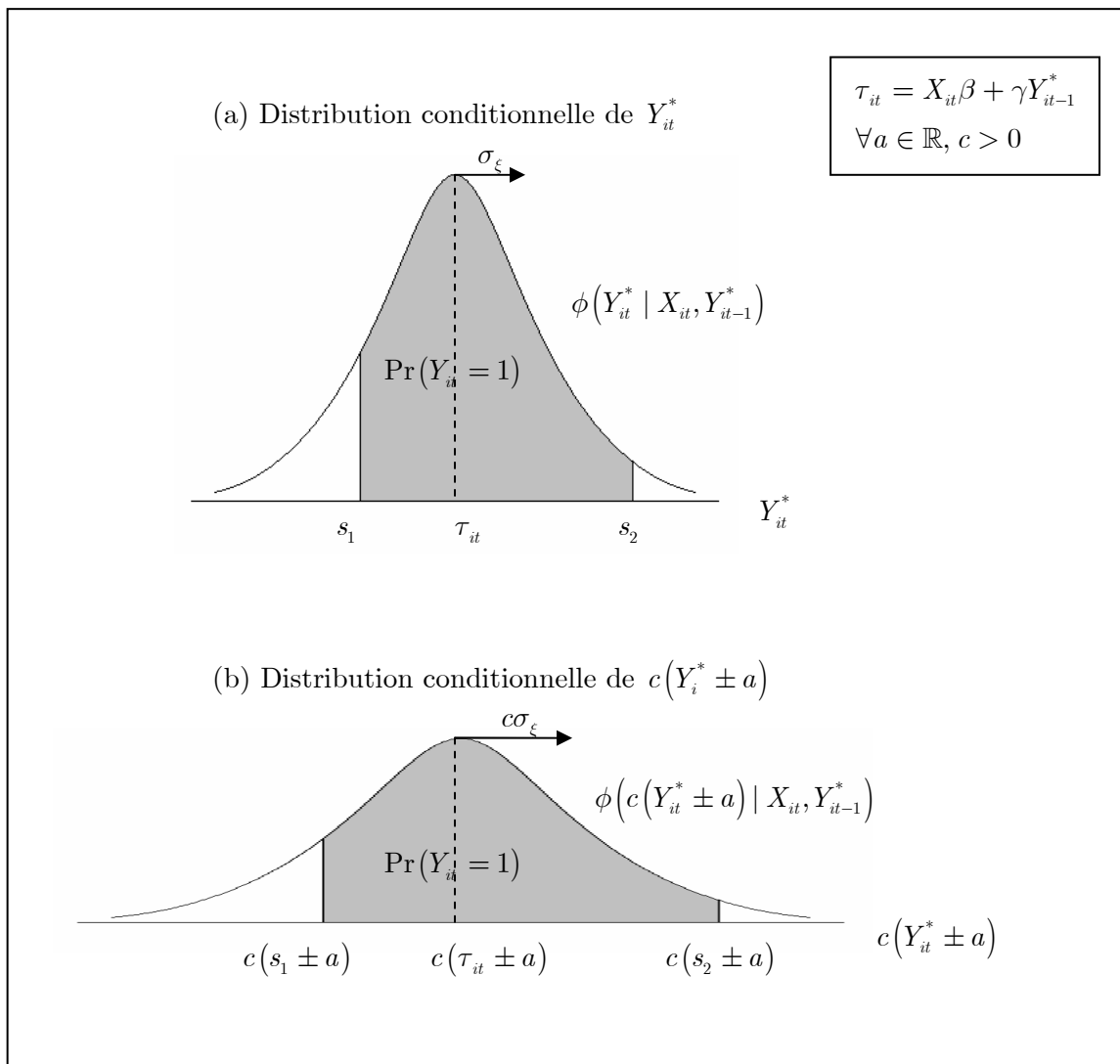
Pour notre analyse, cette condition impose que les variables latentes soient négatives pour les ménages non-motorisés, et inversement, positives pour les ménages motorisés : ceci nous facilite la mise en lumière de comportement d'habitude de motorisation si l'estimation révèle un coefficient γ significativement positif, où autrement dit, si l'état de (non) motorisation passé encourage l'état de (non) motorisation courant.

En revenant à la figure 3.1-(a), on peut observer l'effet du changement d'échelle suivant : la multiplication des seuils $\{s_1, s_2\}$, de l'espérance τ_{it} , et de l'écart-type σ_ξ par un même facteur $c > 1$ (resp. $0 < c < 1$) entraîne une dilatation (resp. une contractation) horizontale de la densité qui laisse inchangée la probabilité de mono-équipement (figure 3.1-(b)). Sachant la précédente condition d'identification (3.9), la probabilité de mono-équipement de l'agent est donnée par :

$\Pr(Y_{it} = 1) = \int_0^{c \times s_2} \phi(cY_{it}^* | X_{it}, Y_{it-1}^*) d(cY_{it}^*)$. Une seconde condition d'identification consiste à fixer un autre paramètre : par convention dans les modèles probit, ce problème d'échelle est résolu en fixant la variance à un, ici :

$$\sigma_\xi^2 = c^2 \sigma_\xi^2 = 1 \quad (3.10)$$

FIGURE 3.1 : Illustration graphique des problèmes d'identification du modèle probit ordonné



3.2.4 Traitement des conditions initiales

La présence parmi les variables explicatives de l'endogène décalée Y_{it-1}^* est la source dynamique de notre spécification (3.5). Cela pose le problème de la période initiale puisque si t varie de la période 1 (pour l'année 1999) jusque 3 (pour l'année 2001) dans notre cas d'étude, alors la variable latente n'est définie que pour $t \geq 2$ (pour les années 2000 et 2001).

Nous nous inspirons ici d'une approche proposée par HECKMAN (1981), qui suggère d'adopter une forme flexible réduite pour approximer les conditions de départ en $t = 1$ et amorcer le modèle général (3.5). On retrouve également cette idée dans l'estimateur de BLUNDELL et SMITH (SEVESTRE, 2003, p. 130) pour les modèles linéaires dynamiques. Cette forme réduite comprend une spécification « raccourcie » et nécessairement statique, donc exprimée en termes de regressseurs initiaux :

$$Y_{i1}^* = X_{i1}\beta_0 + \varepsilon_{i1}. \quad (3.11)$$

Les variables explicatives X_{i1} répondent ici au même descriptif de l'observation i que X_{it} dans le modèle général, β_0 est ici le vecteur de paramètres initiaux, alors que le terme d'erreur noté ε_{i1} est supposé suivre une $\mathcal{N}(0, \sigma_{\varepsilon_0}^2)$. Par analogie avec la règle du modèle général (3.2), le niveau de motorisation initial est déterminé par la discrétisation de la variable Y_{i1}^* :

$$\begin{aligned} Y_{i1} &= 0 \text{ si } -\infty < Y_{i1}^* \leq s'_1, \\ Y_{i1} &= 1 \text{ si } s'_1 < Y_{i1}^* \leq s'_2, \\ Y_{i1} &= 2 \text{ si } s'_2 < Y_{i1}^* \leq +\infty. \end{aligned} \quad (3.12)$$

Si Y_{i1}^* doit intervenir comme endogène retardée dans le modèle général écrit pour Y_{i2}^* , alors il doit être de même nature qu'une endogène du modèle principal et ainsi obéir aux mêmes clés de passage vers l'état observé de motorisation. Ainsi, les seuils de la règle (3.12) sont donnés par :

$$s'_1 = s_1 ; s'_2 = s_2. \quad (3.13)$$

L'importation des deux seuils $\{s_1, s_2\}$ du modèle général vers le modèle initial constitue deux modalités d'identification suffisantes pour le modèle initial. Il

n'est donc pas nécessaire d'imposer une contrainte sur la variance $\sigma_{\varepsilon_0}^2$, comme c'était le cas dans le modèle général pour σ_ξ^2 .

Par ailleurs, le terme d'hétérogénéité α_i n'apparaît pas explicitement dans le modèle initial : s'il existe, alors celui-ci est absorbé par l'erreur ε_{i1} . Pour en tenir compte, nous introduisons le schéma de corrélation suivant :

$$\begin{aligned} \alpha_i &= \delta\varepsilon_{i1} + v_i, \\ \text{avec : } v_i &\perp \varepsilon_{i1} ; v_i \sim \mathcal{N}(0, \sigma_v^2). \end{aligned} \tag{3.14}$$

3.2.5 Vraisemblance du modèle

Les étapes d'identification et de paramétrage effectuées d'une part, et les conditions initiales mises en place d'autre part, nous pouvons écrire la fonction de vraisemblance de notre modèle. Dans ce but, nous considérons tout d'abord l'agent i , dont la pondération dans le panel est notée w_i (avec $\sum_i w_i = W$), et faisons tout d'abord comme si nous avions la connaissance exacte de toutes ses variables inobservables, exceptés les termes d'erreurs $\{\varepsilon_{i3}, \varepsilon_{i2}, \varepsilon_{i1}\}$. Nous pouvons ainsi calculer L_{i3}^* , L_{i2}^* , L_{i1}^* qui sont respectivement les vraisemblances périodiques des variables latentes Y_{i3}^* , Y_{i2}^* , Y_{i1}^* . En observant les regroupements de paramètres $\Theta = \{\beta, \gamma, \delta, \beta_0, \sigma_\varepsilon^2, \sigma_v^2, \sigma_{\varepsilon_0}^2\}$ et $S = \{s'_1, s'_2, s_1, s_2\}$, nous avons :

$$\begin{aligned} L_{i3}^* &= L(Y_{i3}^* | \Theta, Y_{i2}^*, Y_{i1}^*, X_i, v_i) = \\ &= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} (Y_{i3}^* - X_{i3}\beta - \gamma Y_{i2}^* - \delta(Y_{i1}^* - X_{i1}\beta_0) - v_i)^2\right), \end{aligned} \tag{3.15}$$

$$\begin{aligned} L_{i2}^* &= L(Y_{i2}^* | \Theta, Y_{i1}^*, X_i, v_i) = \\ &= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} (Y_{i2}^* - X_{i2}\beta - \gamma Y_{i1}^* - \delta(Y_{i1}^* - X_{i1}\beta_0) - v_i)^2\right), \end{aligned} \tag{3.16}$$

$$L_{i1}^* = L(Y_{i1}^* | \Theta, X_i, v_i) = \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_0}^2}} \exp\left(-\frac{1}{2\sigma_{\varepsilon_0}^2} (Y_{i1}^* - X_{i1}\beta_0)^2\right). \tag{3.17}$$

Par application de la règle de Bayes, le produit de ces trois quantités (3.15), (3.16) et (3.17) calcule L_i^* la vraisemblance jointe des variables latentes de l'agent i :

$$L_i^* = L(Y_i^* | \Theta, X_i, v_i) = \prod_{t=1}^3 L_{it}^* . \quad (3.18)$$

Le produit pondéré sur l'ensemble des agents de L_i^* nous donne la vraisemblance des variables latentes de notre échantillon (une statistique cruciale pour la suite) :

$$L(Y^* | \Theta, X, v) = \prod_i (L_i^*)^{w_i} . \quad (3.19)$$

En retournant au niveau individuel, et sachant les hypothèses (3.14), on calcule pour i la vraisemblance jointe de ses variables latentes et de v_i par :

$$L(Y_i^*, v_i | \Theta, X_i) = L_i^* \times P(v_i) , \quad (3.20)$$

où $P(v_i)$ est la fonction de densité d'une $\mathcal{N}(0, \sigma_v^2)$:

$$P(v_i) = (2\pi\sigma_v^2)^{-0.5} \exp\left(-0.5\sigma_v^{-2}(v_i - 0)^2\right) .$$

L'intégration de (3.20) sur le domaine des possibles des variables latentes tout d'abord (défini par les systèmes de décisions (3.2) et (3.12) et les seuils S), puis sur v_i ensuite, nous donne la vraisemblance de la motorisation observée de l'observation i :

$$L_i = L(Y_i | \Theta, S, X_i) = \int_{-\infty}^{+\infty} \int_{a_{i1}}^{b_{i1}} \int_{a_{i2}}^{b_{i2}} \int_{a_{i3}}^{b_{i3}} L(Y_i^*, v_i | \Theta, S, X_i) dY_{i3}^* dY_{i2}^* dY_{i1}^* dv_i , \quad (3.21)$$

où les bornes d'intégration sont données par :

$$\left. \begin{array}{l} a_{it} = -\infty \\ b_{it} = \tilde{s}_1 \end{array} \right\} \text{si } Y_{it} = 0 ; \left. \begin{array}{l} a_{it} = \tilde{s}_1 \\ b_{it} = \tilde{s}_2 \end{array} \right\} \text{si } Y_{it} = 1 ; \left. \begin{array}{l} a_{it} = \tilde{s}_2 \\ b_{it} = +\infty \end{array} \right\} \text{si } Y_{it} = 2;$$

avec : $\tilde{s}_1 = s'_1$ pour $t = 1$; $\tilde{s}_1 = s_1$ pour $t = \{2, 3\}$;

$\tilde{s}_2 = s'_2$ pour $t = 1$; $\tilde{s}_2 = s_2$ pour $t = \{2, 3\}$,

(3.22)

et où le jeu de contraintes d'identification est :

$$s_1 = 0 ; s'_1 = s_1 ; s'_2 = s_2 ; \sigma_\varepsilon^2 = 1 - (\delta^2 \sigma_{\varepsilon_0}^2 + \sigma_v^2).$$
(3.23)

Par le produit pondéré des vraisemblances individuelles en (3.21), nous obtenons la vraisemblance des états de motorisation de notre échantillon :

$$L(Y | \Theta, S, X) = \prod_i (L_i)^{w_i}.$$
(3.24)

La non connaissance explicite des variables latentes Y^* a pour effet d'introduire des intégrales emboîtées dans la fonction (3.21), ce qui rend l'estimation des paramètres par maximum de vraisemblance très complexe. Si toutes les variables latentes Y_{it}^* étaient connues, la vraisemblance (3.21) ne contiendrait plus qu'une intégrale sur v_i , ce qui est beaucoup plus gérable, et un estimateur traditionnel pour modèles linéaires dynamiques avec erreurs composées aurait pu alors être envisagé.

C'est pourquoi une procédure d'estimation par simulation des variables inobservées peut considérablement nous simplifier la tâche. En particulier, nous savons depuis les travaux d'ALBERT et CHIB (1993) que les méthodes de Monte Carlo par chaînes de Markov (MCMC) tel que l'échantillonnage de Gibbs avec augmentation de données sont particulièrement adaptées et efficaces pour l'estimation de modèles économétriques à base de variables latentes. Ces méthodes s'accommodent si bien de la statistique bayésienne qu'elles ont été le moteur de sa résurrection au cours des dix dernières années. Même si ces méthodes MCMC ne lui sont pas exclusives (TANNER, 1991), nous adoptons un cadre d'estimation bayésien pour traiter notre modèle ordonné du niveau de motorisation des ménages.

En choisissant la spécification (3.5) pour le modèle latent, nous souhaitons mettre les propriétés dynamiques avantageuses d'une spécification linéaire autorégressive basique, avec ses outils d'analyse, au service de notre modèle de choix ordonnés. Cette méthodologie nous a été inspirée par les travaux relatifs à

l'estimation du modèle probit multinomial dynamique de PAAP et FRANCES (2000). Dans la littérature des transports, il existe quelques travaux de modélisation dynamique portant sur l'état de motorisation d'observations désagrégées. Parmi ceux cités en introduction, aucune étude à notre connaissance n'utilise une spécification latente autorégressive, la raison tenant à la nature des termes dynamiques utilisés. Toutes en effet utilisent au mieux un ensemble d'indicatrices du niveau (observé) de motorisation passée comme facteurs explicatifs. Dans la pratique, il est évidemment plus commode d'estimer les paramètres d'une spécification lorsque l'ensemble des régresseurs est connu, plutôt que d'utiliser une spécification incorporant de telles variables latentes non observées comme dans notre cas. Hormis quelques conclusions attestant du rôle significatif et positif de la motorisation passée sur la motorisation courante, l'interprétation et l'exploitation des résultats dynamiques dans ces études ne pouvaient, semble-t-il, pas être aisément approfondies. Notre modèle latent autorégressif (3.5) permet des conclusions dynamiques plus intéressantes. Notamment, nous verrons ultérieurement que son estimation nous donne la capacité de mesurer le temps nécessaire aux agents d'ajuster leur décision de motorisation aux changements de caractéristiques, mais aussi d'évaluer individuellement les variations/élasticités de probabilités à court et long terme des trois états de motorisation que ces changements induisent.

3.3 L'interprétation du modèle

Selon le modèle probit ordonné, la modalité observée d'un ménage traduit que sa variable latente se situe dans la fourchette correspondante. Un modèle dont la spécification retenue aurait été statique, qui n'aurait fait dépendre la variable latente que de variables contemporaines, impliquerait un ajustement instantané total aux variations de ses déterminants. Autrement dit, une variation de revenu ou un changement de résidence d'un ménage par exemple serait complètement et immédiatement répercuté sur le niveau de sa variable latente, et donc sur son niveau de motorisation. On se doute évidemment qu'une telle flexibilité n'est que théorique, surtout concernant l'équipement automobile. En effet, l'acquisition et la cession d'une automobile par un agent sont des actes généralement précédés d'étapes et d'activités coûteuses en temps, les plus évidentes étant la réflexion, la recherche (d'informations, de prix, des modalités de stationnement et

d'assurances...), voire l'attente (d'une meilleure opportunité, de conception du modèle désiré, ou l'hésitation). Il est donc pratiquement acquis que la décision individuelle de motorisation revêt une dimension temporelle et repose sur une spécification dynamique de la variable latente. Même si ce n'était pas le cas, notre spécification (3.5) n'en serait pas pour autant inopérante puisque le cas statique lui est un cas particulier, pour $\gamma = 0$.

En substituant les endogènes retardées par leur expression, on peut réécrire le processus générateur de la variable latente (3.5) et le faire apparaître comme une réponse « lissée » aux modifications abruptes des déterminants X , qui nous servira pour les calculs d'effets marginaux de court et long terme :

$$\begin{aligned} Y_{it+h}^* &= \gamma^t Y_{ih}^* + \sum_{s=0}^{t-1} \gamma^s (X_{it+h-s} \beta + \alpha_i + \varepsilon_{it+h-s}) \\ &= \gamma^h Y_{it}^* + \sum_{s=0}^{h-1} \gamma^s (X_{it+h-s} \beta + \alpha_i + \varepsilon_{it+h-s}). \end{aligned} \quad (3.25)$$

Puisque les termes aléatoires ε_{iz} sont normaux, homoscedastiques, et d'espérance nulle quel que soit z , alors l'espérance, la variance, et la distribution de Y_{it}^* sachant les déterminants présents et passés $X_{i,\leq t} = \{X_{it}, \dots, X_{i1}\}$, et pour $t \gg 0$, nous sont donnés dans la relation (3.25) par :

$$\begin{aligned} \mu_{it} &= E[Y_{it}^* | X_{i,\leq t}] \approx \sum_{s=0}^{t-1} \gamma^s (X_{it-s} \beta), \\ \omega_t &= V[Y_{it}^* | X_{i,\leq t}] = V\left(\sum_{s=0}^{t-1} \gamma^s (\alpha_i + \varepsilon_{it-s})\right) \approx \left(\frac{1}{1-\gamma}\right)^2 \sigma_\alpha^2 + \frac{1}{1-\gamma^2} \sigma_\varepsilon^2; \\ (Y_{it}^* | X_{i,\leq t}) &\sim \mathcal{N}(\mu_{it}, \omega_t), \text{ pour } (t \gg 0 : \gamma^t \approx 0). \end{aligned} \quad (3.26)$$

Dans la partie suivante, nous examinons les propriétés dynamiques des statistiques en (3.26), et mesurons l'effet de court et long terme qu'induit une modification d'un déterminant sur la variable latente avec les multiplicateurs linéaires standards.

3.3.1 Effets latents et multiplicateurs

3.3.1.1 Dans le court terme

La relation (3.25) établit que l'impact d'une modification de la k -ième caractéristique en $t - s$ sur la variable latente courante se calcule par :

$$\frac{dY_{it}^*}{dX_{ikt-s}} = \gamma^s \beta_k. \quad (3.27)$$

Nous retrouvons un résultat illustrant cette idée ordinaire que l'influence sur le présent des événements passés diminue avec le temps. En effet, pour $0 \leq \gamma < 1$, nous obtenons que l'impact absolu est décroissant avec le temps s qui les sépare :

$$\lim_{s \rightarrow \infty} \left(\left| \frac{dY_{it}^*}{dX_{ikt-s}} \right| \right) = 0. \quad (3.28)$$

L'effet de court terme sur la variable latente Y_{it}^* correspond à l'effet instantané d'une variation de caractéristique X_{ikt} , et s'obtient à partir de (3.27) pour $s = 0$:

$$dY_{it}^* = m_k^{\text{CT}} \times dX_{ikt}, \quad (3.29)$$

où $m_k^{\text{CT}} = \beta_k$ est le multiplicateur de court terme de X_{ik} sur la variable latente.

3.3.1.2 Dans le long terme

Nous supposons que la période t est suffisamment ancienne pour que ses effets sur la variable latente en $t + h$ soient dissous, ou formellement que h est suffisamment grand pour que $\gamma^h \approx 0$ dans l'équation (3.25). Si de plus nous considérons le gel des caractéristiques tel que $X_{i,[t;t+h]} = \{X_{it}, \dots, X_{it+h}\} = \underline{X}_i$, alors (3.25) se simplifie, et les moments limites de Y_{it+h}^* sont donnés par :

$$Y_{it+h}^* = \sum_{s=0}^{h-1} \gamma^s (\underline{X}_i \beta + \alpha_i + \varepsilon_{it+h-s}), \quad (3.30)$$

$$\begin{aligned}\mu_{(\underline{X}_i)}^{\text{LT}} &= \lim_{h \rightarrow \infty} \mathbb{E}[Y_{it+h}^* \mid X_{i,[t;t+h]} = \underline{X}_i] = \frac{1}{1-\gamma} \underline{X}_i \beta, \\ \omega^{\text{LT}} &= \lim_{h \rightarrow \infty} \mathbb{V}[Y_{it+h}^* \mid X_{i,[t;t+h]} = \underline{X}_i] = \frac{1}{(1-\gamma)^2} \sigma_\alpha^2 + \frac{1}{1-\gamma^2} \sigma_\varepsilon^2, \\ (Y_{i\infty}^* \mid X_{i,\geq t} = \underline{X}_i, \beta, \alpha_i) &\sim \mathcal{N}(\mu_{(\underline{X}_i)}^{\text{LT}}, \omega^{\text{LT}}),\end{aligned}\tag{3.31}$$

et en remarquant notamment que $\omega^{\text{LT}} \approx \omega_t$. Nous constatons dans ces conditions que les statistiques caractérisant la distribution de Y_{it+h}^* en (3.31) sont indépendantes des périodes : celle-ci est donc stationnaire et l'équilibre latent de long terme est atteint. Le multiplicateur de long terme, ou d'équilibre (GREENE, 2003, p.562) associé à la k -ième variable explicative de \underline{X}_i nous est donné par :

$$m_k^{\text{LT}} = \lim_{h \rightarrow \infty} \left(\frac{dY_{it+h}^*}{d\underline{X}_{ik}} \right) = \frac{\beta_k}{1-\gamma},\tag{3.32}$$

L'effet latent de long terme résultant d'un changement permanent de la caractéristique k correspond à la variation induite sur $Y_{i\infty}^*$:

$$dY_{i\infty}^* = d\underline{X}_{ik} \times m_k^{\text{LT}}.\tag{3.33}$$

Notons qu'une autre manière d'aboutir aux multiplicateurs de long terme (3.32) est de cumuler les multiplicateurs de court terme et décalés de l'équation (3.27) pour $s > 0$: $m_k^{\text{LT}} = \beta_k + \sum_{s=1}^{\infty} \gamma^s \beta_k = (1-\gamma)^{-1} \beta_k$.

Les effets latents de court et long terme que nous venons de dériver ne sont que des étapes intermédiaires : ils interviennent au niveau de la variable latente Y_i^* , et n'entretiennent ainsi qu'un rapport indirect avec les états de motorisation Y_i . Il est donc nécessaire d'étendre l'analyse aux effets marginaux de court et long terme sur les probabilités des états de motorisation.

3.3.2 Effets marginaux à court et long terme

3.3.2.1 Dans le court terme

Les effets marginaux sont l'outil d'interprétation habituel du modèle probit ordonné, et évaluent dans notre étude les variations de probabilité $\Pr[Y_{it} = j \mid X_{i,\leq t}]$ induites par une modification de caractéristiques à la période t .

Une façon pratique de calculer ces probabilités est de « centrer-réduire » la distribution de $Y_{it}^* \mid X_{i,\leq t}$ et d'utiliser la fonction de répartition Φ d'une $\mathcal{N}(0,1)$:

$$\begin{aligned}\Pr[Y_{it} = 0 \mid X_{i,\leq t}] &= \Phi\left((s_1 - \mu_{it})/\sqrt{\omega_t}\right), \\ \Pr[Y_{it} = 1 \mid X_{i,\leq t}] &= \Phi\left((s_2 - \mu_{it})/\sqrt{\omega_t}\right) - \Phi\left((s_1 - \mu_{it})/\sqrt{\omega_t}\right), \\ \Pr[Y_{it} = 2 \mid X_{i,\leq t}] &= 1 - \Phi\left((s_2 - \mu_{it})/\sqrt{\omega_t}\right).\end{aligned}\tag{3.34}$$

L'effet à court terme d'une variation de X_{it} , notée ΔX_{it} , est une variation de μ_{it} qui fait intervenir les multiplicateurs de court terme dans (3.29), et mesurée par $\Delta\mu_{it} = \Delta X_{it} m^{\text{CT}}$. Il résulte de cette variation un nouveau système de probabilité :

$$\begin{aligned}\Pr[Y_{it} = 0 \mid X_{i,\leq t}, \Delta X_{it}] &= \Phi\left(\frac{s_1 - (\mu_{it} + \Delta X_{it} m^{\text{CT}})}{\sqrt{\omega_t}}\right), \\ \Pr[Y_{it} = 1 \mid X_{i,\leq t}, \Delta X_{it}] &= \Phi\left(\frac{s_2 - (\mu_{it} + \Delta X_{it} m^{\text{CT}})}{\sqrt{\omega_t}}\right) - \Phi\left(\frac{s_1 - (\mu_{it} + \Delta X_{it} m^{\text{CT}})}{\sqrt{\omega_t}}\right), \\ \Pr[Y_{it} = 2 \mid X_{i,\leq t}, \Delta X_{it}] &= 1 - \Phi\left(\frac{s_2 - (\mu_{it} + \Delta X_{it} m^{\text{CT}})}{\sqrt{\omega_t}}\right).\end{aligned}\tag{3.35}$$

Les effets marginaux de court terme, notés $\Lambda_{ij}^{\text{CT}}(\Delta X_{it})$ se déduisent en différenciant les systèmes de probabilités (3.35) et (3.34), et s'interprètent pour chaque état de motorisation $Y_{it} = j$ comme la variation de probabilité résultant d'un changement de caractéristique ΔX_{it} :

$$\Lambda_{ij}^{\text{CT}}(\Delta X_{it}) = \Pr[Y_{it} = j \mid X_{i,\leq t}, \Delta X_{it}] - \Pr[Y_{it} = j \mid X_{i,\leq t}].\tag{3.36}$$

3.3.2.2 Dans le long terme

Dans une démarche identique, la conséquence à long terme d'une variation permanente de \underline{X}_i mesurée par $\Delta \underline{X}_i$ sur les états de motorisation est le système de probabilités suivant, qui fait intervenir les multiplicateurs d'équilibre m^{LT} de (3.33):

$$\begin{aligned}
 \Pr[Y_{i\infty}=0 \mid X_{i,\geq t}=\underline{X}_i, \Delta\underline{X}_i] &= \Phi\left(\frac{s_1 - (\mu_{(\underline{X}_i)}^{\text{LT}} + \Delta\underline{X}_i m^{\text{LT}})}{\sqrt{\omega^{\text{LT}}}}\right), \\
 \Pr[Y_{i\infty}=1 \mid X_{i,\geq t}=\underline{X}_i, \Delta\underline{X}_i] &= \Phi\left(\frac{s_2 - (\mu_{(\underline{X}_i)}^{\text{LT}} + \Delta\underline{X}_i m^{\text{LT}})}{\sqrt{\omega^{\text{LT}}}}\right) - \Phi\left(\frac{s_1 - (\mu_{(\underline{X}_i)}^{\text{LT}} + \Delta\underline{X}_i m^{\text{LT}})}{\sqrt{\omega^{\text{LT}}}}\right), \\
 \Pr[Y_{i\infty}=2 \mid X_{i,\geq t}=\underline{X}_i, \Delta\underline{X}_i] &= 1 - \Phi\left(\frac{s_2 - (\mu_{(\underline{X}_i)}^{\text{LT}} + \Delta\underline{X}_i m^{\text{LT}})}{\sqrt{\omega^{\text{LT}}}}\right).
 \end{aligned} \tag{3.37}$$

En différenciant les systèmes de probabilités (3.37) et (3.34), nous obtenons les effets marginaux de long terme $\Lambda_{ij}^{\text{LT}}(\Delta\underline{X}_i)$, qui mesurent les variations limites de probabilité résultant d'un changement de caractéristiques permanent $\Delta\underline{X}_i$:

$$\Lambda_{ij}^{\text{LT}}(\Delta\underline{X}_i) = \Pr[Y_{i\infty}=j \mid X_{i,>t}=\underline{X}_i, \Delta\underline{X}_i] - \Pr[Y_{it}=j \mid X_{i,\leq t}]. \tag{3.38}$$

3.4 Estimation bayésienne MCMC

Les techniques d'estimation à l'aide de méthodes de Monte Carlo par chaînes de Markov (MCMC) constituent un apport majeur pour l'analyse des données. Pour JACKMAN (2000), elles sont « *probablement le développement statistique le plus excitant* » depuis les années 1990. Ces techniques viennent notamment secourir des méthodes bayésiennes jusqu'alors peu attractives face aux méthodes classiques, tant leurs manipulations algébriques sont souvent lourdes et décourageantes⁶⁶. La réunion du cadre bayésien et de l'outil MCMC rend désormais accessible l'estimation de modèles traditionnellement complexes, et autorise des spécifications larges pour une bonne exploitation des données de panel notamment. Dans la littérature, ALBERT et CHIB (1993) font figure de pionniers dans l'usage de ces méthodes pour l'estimation de modèles probit. Notamment sur ces bases, les travaux de McCULLOCH et ROSSI (1994) ont exposé la première estimation bayésienne du modèle probit multinomial. Par la suite, ce modèle a été exploité par NOBILE et al. (1997) pour modéliser le degré de motorisation aux Pays-Bas, en utilisant un algorithme MCMC hybride.

⁶⁶ La lecture de l'analyse bayésienne du modèle de régression classique dans GREENE (2003, p. 430), est en ce sens assez convaincante.

3.4.1 Quelques notions bayésiennes d'économétrie

La pièce maîtresse de l'analyse bayésienne est la célèbre règle de Bayes en personne : $P(A | B) = (P(B | A) \times P(A)) / P(B)$. Elle fait ici apparaître deux types de probabilités : marginales $(P(A), P(B))$ d'une part, puis conditionnelles $(P(A | B), P(B | A))$ d'autre part. La lecture usuelle du produit de cette règle interprète $P(A | B)$, comme la probabilité de voir se réaliser l'évènement A sachant que l'évènement B s'est réalisé. L'adaptation économétrique de la règle de Bayes constitue le cœur de l'économétrie bayésienne. En général, l'économètre porte un grand intérêt à l'estimation de paramètres θ et utilise pour cela des données y à sa disposition comme source d'information. Autrement dit, l'estimation traite de l'état des connaissances concernant θ après l'examen de (ou « *sachant* ») y , et que l'on note $P(\theta | y)$: telle est la préoccupation centrale du chercheur bayésien. En décomposant avec la règle de Bayes précédente, nous obtenons :

$$P(\theta | y) = \frac{P(y | \theta) \times P(\theta)}{P(y)}. \quad (3.39)$$

Les données y étant constituées d'observations, elles ne sont pas modulables, et leur probabilité $P(y)$ est donc constante quelque soit la valeur des paramètres θ . Autrement dit, puisque l'intérêt ne porte que sur θ , qui n'intervient pas dans $P(y)$, ce dernier terme peut être ignoré. Dans l'usage, les données y ont une valeur informative et servent l'actualisation des paramètres. Ainsi il vient que :

$$P(\theta | y) \propto P(y | \theta) \times P(\theta). \quad (3.40)$$

Dans la terminologie bayésienne, l'élément $P(\theta | y)$ renvoie à la distribution ou densité de probabilité des paramètres *a posteriori*, le terme $P(y | \theta)$ est la fonction de densité des observations sachant les paramètres, alors que $P(\theta)$ concerne la distribution ou fonction de densité des paramètres *a priori*.

La densité *a priori* des paramètres, $P(\theta)$, traduit les croyances initiales du modélisateur, ou toute l'information qu'il détient avant l'exploration des données. Elle peut être ajustée pour en refléter diverses nuances : si aucune information n'est *a priori* disponible sur la valeur des paramètres, alors il est usuel et pratique

d'assigner à θ une distribution *a priori* $P(\theta)$ conjuguée⁶⁷ et très diffuse, c'est-à-dire caractérisée par une variance élevée et une allure (quasiment) « plate ». Au contraire, s'il existe de l'information statistique concernant θ dans la littérature, provenant par exemple de travaux antérieurs ayant utilisé une autre source de données, alors le modélisateur pourra s'en servir pour spécifier une distribution $P(\theta)$ informative.

La fonction de densité des observations sachant les paramètres, notée $P(y | \theta)$, n'est autre que la fonction de vraisemblance classique : elle contient l'information issue des données y sur les paramètres θ . A cet égard, et pour adopter une notation plus usuelle, nous la noterons $L(y | \theta)$:

$$P(y | \theta) \equiv L(y | \theta). \quad (3.41)$$

La distribution *a posteriori* des paramètres est selon la relation (3.40) proportionnelle au produit de la fonction de vraisemblance et de la distribution *a priori* des paramètres. Cette distribution, que le bayésien cherche à caractériser, focalise l'intérêt puisqu'elle condense toute l'information - celle *a priori* et celle contenue dans les données - concernant les paramètres θ .

Le passage de la distribution *a priori* vers celle *a posteriori* peut être vu comme un processus bayésien d'apprentissage, où la mise à jour des connaissances initiales s'effectue à lumière des données. D'une façon imagée, la vraisemblance agit telle une boîte, qui reçoit en input la distribution *a priori* $P(\theta)$, la modifie selon l'information dont elle dispose en y , et renvoie en output la distribution *a posteriori*, $P(\theta | y)$.

Ces concepts bayésiens maintenant définis, il est important de noter que l'usage de distributions *a priori* très diffuses revient à approximer $P(\theta)$ par une constante. Il en ressort que $P(\theta | y)$ devient proportionnelle à $P(y | \theta)$ et que la vraisemblance décide seule de l'allure de la distribution *a posteriori*. Dans ce cas d'*a priori*, dits « non informatifs », les méthodes classiques ou bayésiennes donnent lieu à des inférences sur θ équivalentes.

Cependant, la densité *a posteriori* des paramètres $P(\theta | y)$ n'est caractérisable par ses moments qu'au prix de résolutions d'intégrales, et de

⁶⁷ Une distribution *a priori* est dite « conjuguée » lorsque son produit avec la vraisemblance donne naissance à une distribution *a posteriori* de même nature que la distribution *a priori*. Par exemple, le produit d'une vraisemblance avec une distribution *a priori* normale conjuguée sera proportionnel à une distribution *a posteriori* normale.

manipulations algébriques fort décourageantes. Cela explique en particulier pourquoi l'économétrie bayésienne a été si longtemps laissée en disgrâce face aux méthodes classiques, beaucoup plus accessibles pour une large majorité de modèles (cf. note 66). Fort heureusement, et grâce aux propriétés de l'échantillonneur de Gibbs, il est possible de reconstituer empiriquement la forme de la fonction de densité *a posteriori* par des échantillonnages successifs dans les lois *a posteriori* conditionnelles des paramètres, et de calculer les moments de leurs distributions marginales.

3.4.2 Une méthode MCMC : l'échantillonneur de Gibbs⁶⁸

L'échantillonneur de Gibbs est une méthode de Monte Carlo par chaîne de Markov qui nous vient de la science physique, dont on attribue généralement la paternité à GEMAN et GEMAN (1984), et l'introduction dans le domaine statistique à GELFAND et SMITH (1990). A l'origine, Geman et Geman ont utilisé cet outil pour la restauration d'images. Les auteurs ont été inspiré par un résultat de BESAG (1974), qui montre l'équivalence de détenir une distribution jointe de $m \geq 2$ variables aléatoires, ou détenir m distributions conditionnelles aux autres $m - 1$ variables. Dans leur contexte (l'imagerie), la distribution jointe des valeurs d'un ensemble de m pixels est compliquée, car de dimension très élevée. Au contraire, les distributions conditionnelles de chaque pixel sont nettement plus manipulables, et sont des distributions de Gibbs. C'est pourquoi l'algorithme mis au point par Geman et Geman porte le nom d'échantillonneur de Gibbs, bien que leur technique puisse être appliquée dans un autre contexte avec d'autres types de distributions conditionnelles.

3.4.2.1 Fonctionnement

L'algorithme d'échantillonnage de Gibbs est une méthode de Monte Carlo par chaîne de Markov (MCMC) qui constitue un cas particulier de l'algorithme, également MCMC, de Metropolis-Hastings (ROBERT, 1996, p.166). Concrètement, il permet la simulation d'une distribution jointe à partir de tirages dans ses distributions conditionnelles. Lorsque le vecteur de paramètres θ peut être

⁶⁸ En l'honneur de Josiah Willard Gibbs (1839-1903), physicien américain de grande renommée pour ses travaux en thermodynamique et en statistique mécanique.

partitionné en $(\theta_1, \theta_2, \theta_3, \dots, \theta_K)$, et d'une manière telle que l'on puisse faire des tirages dans les distributions *a posteriori* conditionnelles suivantes :

$$P(\theta_k | y, \theta_{-k}) \text{ pour tout } k ,$$

avec $\theta_{-k} = \{\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_K\}$, alors des tirages successifs dans les lois conditionnelles, continûment actualisées par les paramètres « fraîchement » échantillonnés forment asymptotiquement des tirages aléatoires dans la distribution jointe *a posteriori* des paramètres $P(\theta | y)$. De cette façon, l'échantillonneur de Gibbs simule θ_k à partir des ces distributions conditionnelles, récursivement, et en mettant systématiquement à jour les distributions conditionnelles par la valeur du dernier paramètre échantillonné (CHIB, 1995). C'est-à-dire que l'on effectue une boucle avec les tirages suivants :

$$\begin{aligned} 1 : & \quad \theta_1^{(g+1)} \text{ dans } P(\theta_1 | y, \theta_2^{(g)}, \theta_3^{(g)}, \dots, \theta_K^{(g)}), \\ 2 : & \quad \theta_2^{(g+1)} \text{ dans } P(\theta_2 | y, \theta_1^{(g+1)}, \theta_3^{(g)}, \dots, \theta_K^{(g)}), \\ 3 : & \quad \theta_3^{(g+1)} \text{ dans } P(\theta_3 | y, \theta_1^{(g+1)}, \theta_2^{(g+1)}, \dots, \theta_K^{(g)}), \\ \vdots : & \quad \vdots \quad \quad \quad \vdots \\ K : & \quad \theta_K^{(g+1)} \text{ dans } P(\theta_K | y, \theta_1^{(g+1)}, \theta_2^{(g+1)}, \dots, \theta_{K-1}^{(g+1)}), \\ K+1 : & \quad g \leftarrow g + 1 ; \text{ retour en 1,} \end{aligned} \tag{3.42}$$

où $\theta_1^{(g+1)}$ est le $(g+1)$ ième échantillonnage de θ_1 , etc... Ce processus génère une chaîne de Markov (c'est la partie « Markov Chain » du sigle MCMC) dont les éléments convergent en distribution vers des tirages de $P(\theta | y)$, la fonction de densité jointe des paramètres *a posteriori*. La convergence atteinte, la distribution de la chaîne de Markov est alors dite stationnaire sur $P(\theta | y)$ (TIERNEY, 1994). L'avantage de détenir *in fine* un échantillon empirique de distribution stationnaire $P(\theta | y)$ à la place d'une fonction de densité est qu'il permet de calculer très aisément toutes les statistiques (moments, écarts-types, quantiles...) la caractérisant. Il n'est plus nécessaire de calculer le moment postérieur de $z(\theta)$ par⁶⁹ :

$$\tilde{H} = \int_{\theta} z(\theta) \times P(\theta | y) d\theta ,$$

⁶⁹ z est une simple fonction de θ . Par exemple, pour $z(\theta) = \theta$, nous avons la formulation de l'espérance.

puisque la moyenne de la fonction z des éléments de la chaîne de Markov converge (presque sûrement) vers l'espérance de la densité *a posteriori* de $z(\theta)$. C'est la partie « Monte Carlo » de MCMC) :

$$\hat{H} = \frac{1}{G} \sum_{g=1}^G z(\theta^{(g)}) \xrightarrow{p.s.} E(z(\theta) | y).$$

Sous les conditions de Lebesgue et lorsque G est grand, \hat{H} est une estimation consistante de \tilde{H} en vertu du théorème ergodique (STOUT, 1974 ; TIERNEY, 1994 ; ROBERT, 1996).

3.4.2.2 Une extension pratique : l'étape « augmentation data »

Il s'agit d'une étape astucieuse que l'on doit à TANNER et WONG (1987) qui consiste à augmenter l'espace des paramètres avec le vecteur de variables latentes Y^* . « Pour les bayésiens, toutes les quantités inobservables peuvent faire l'objet d'inférence sans se soucier qu'elles sont celles de paramètres ou de variables latentes » (ROSSI et ALLENBY, 2000, p. 309). En incluant cette étape dans l'algorithme de Gibbs, on contourne l'évaluation de $\int P(\theta, Y^* | y) dY^*$ pour calculer la distribution marginale *a posteriori* $P(\theta | y)$. L'idée reste simple : si l'on parvient à récupérer de l'échantillonneur de Gibbs une suite de tirages $\{\theta^{(g)}, Y^{*(g)}\}$ issus de la loi jointe $P(\theta, Y^* | y)$, alors il suffit de supprimer l'ensemble des $Y^{*(g)}$ échantillonnés pour marginaliser la distribution des paramètres θ . Formellement, l'échantillonneur de Gibbs avec augmentation de données devient une succession de tirages dans des distributions conditionnelles suivantes :

Etape 1 : Elle correspond aux K premières étapes de l'algorithme (3.42).

Pour $k = 1 \rightarrow K$,

k : $\theta_k^{(g+1)}$ tirés dans $P(\theta_k | \theta_{<k}^{(g+1)}, \theta_{>k}^{(g)}, Y^{*(g)}, y)$

Etape 2 : Augmentation de données :

Pour $i = 1 \rightarrow N$,

$K + i$: $Y_i^{*(g+1)}$ tiré dans $P(Y_i^* | \theta^{(g+1)}, y)$,

$K + N + 1$: $g \leftarrow g + 1$. Retour en l'étape 1.

Si l'étape d'augmentation de données nous affranchit de toute intégrale ou évaluation de probabilité, elle est venue agrandir l'espace d'échantillonnage, ce qui généralement s'accompagne d'une corrélation en série des tirages accrue, et par conséquent d'une convergence plus lente vers la distribution stationnaire $P(\theta | y)$. Pour contrôler la convergence, nous utilisons dans ce chapitre le diagnostic de GEWEKE (1992), dont nous faisons la description rapide dans le prochain paragraphe. Nous expliciterons ensuite les distributions conditionnelles *a posteriori* des paramètres de notre modèle, nécessaires à la bonne marche de l'échantillonnage de Gibbs.

3.4.3 Le diagnostic de convergence de Geweke

GEWEKE (1992) fait remarquer que les chaînes d'échantillonnage générées peuvent être vues sous l'angle des séries chronologiques, et propose un diagnostic fondé sur leur analyse spectrale. L'intuition est qu'après une phase de convergence (ou « *burnin stage* »), la moyenne des premiers éléments d'une chaîne de $\theta_k^{(g)}$ doit être sensiblement la même que celle des derniers éléments. Concrètement, la méthode est la suivante.

Sous l'hypothèse que la chaîne d'un paramètre θ_k a convergé vers sa distribution stationnaire à l'itération n_0 , alors les tirages $\theta_k^{(1)}, \dots, \theta_k^{(n_0)}$ sont écartés, et la chaîne $\theta_k^{(n_0+1)}, \dots, \theta_k^{(n_0+S)}$, de longueur S , est stationnaire. Dans cette situation, l'auteur propose de découper cette chaîne en trois parties : la première (A) concerne la sous-séquence de tirages $\theta_k^{(n_0+1)}, \dots, \theta_k^{(n_A)}$, la seconde (B) est constituée des tirages $\theta_k^{(n_A+1)}, \dots, \theta_k^{(n_B)}$, et la troisième (C) est composée des derniers tirages $\theta_k^{(n_B+1)}, \dots, \theta_k^{(n)}$. Respectivement, les longueurs de ces trois sous-séquences sont données par $S_A = n_A - n_0$; $S_B = n_B - n_A$; $S_C = n - n_B$, avec $S_A + S_B + S_C = S$. Notons les moyennes des sous-séquences (A) et (C) :

$$\bar{\theta}_k^A = \frac{1}{S_A} \sum_{g=n_0+1}^{n_A} \theta_k^{(g)} ; \bar{\theta}_k^C = \frac{1}{S_C} \sum_{g=n_B+1}^n \theta_k^{(g)} ,$$

dont les écarts-types standard estimés respectifs nous sont donnés par :

$$\sqrt{\frac{\hat{S}_{\theta_k}^A(0)}{S_A}} ; \sqrt{\frac{\hat{S}_{\theta_k}^C(0)}{S_C}} ,$$

où $\hat{S}_{\theta_k}^A(0)$ et $\hat{S}_{\theta_k}^C(0)$ sont des estimations consistantes de la densité spectrale évaluée à la fréquence 0 de la première et la dernière sous-séquence respectivement, et dont l'usage est justifié par la dépendance en série des tirages $\theta_k^{(g)}$ (voir RIPLEY, 1987). Pour des ratios S_A/S , S_B/S , S_C/S constants, et à mesure que la longueur totale S augmente, alors la distance S_B qui sépare les sous-séquences (A) et (C) augmente, ce qui tend à les rendre davantage indépendantes. Dès lors, nous pouvons invoquer le théorème de la limite centrale :

$$CD = \frac{\bar{\theta}_k^A - \bar{\theta}_k^C}{\sqrt{\frac{\hat{S}_{\theta_k}^A(0)}{S_A} + \frac{\hat{S}_{\theta_k}^C(0)}{S_C}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Si la statistique $|CD|$ calculée est trop grande ($|CD| > 1.96$), alors l'hypothèse d'égalité des moyennes des sous-séquences (A) et (C) est rejetée, suggérant que la chaîne du paramètre θ_k n'avait pas fini de converger en n_0 . Dans la pratique, GEWEKE recommande l'utilisation des ratios suivants : $S_A/S = 0.1$, $S_C/S = 0.5$. COWLES et CARLIN (1996), ainsi que BROOKS et ROBERTS (1998) (entre autres) font état de quelques critiques de ce diagnostic.

3.4.4 Distributions conditionnelles a posteriori du modèle

Toutes les distributions conditionnelles sont formellement obtenues à partir de la relation de proportionnalité décrite en (3.40). En augmentant l'espace des paramètres avec celui des variables latentes, nous pouvons réécrire cette relation de proportionnalité avec les notations de notre modèle :

$$P(\Theta, S, Y^*, v \mid Y, X) \propto L(Y, Y^* \mid \Theta, S, v, X) \times P(\Theta) \times P(S) \times P(v). \quad (3.43)$$

En opérant une décomposition conditionnelle sur $L(Y, Y^* \mid \Theta, S, v, X)$, l'équation précédente est équivalente à :

$$P(\Theta, S, Y^*, v \mid Y, X) \propto L(Y \mid Y^*, \Theta, S, v, X) L(Y^* \mid \Theta, S, v, X) \times P(\Theta) \times P(S) \times P(v). \quad (3.44)$$

En vertu de la règle décisionnelle décrite en (3.2) et (3.12), les variables latentes Y^* et les seuils $S = \{s_1, s_2, s'_1, s'_2\}$ sont des éléments dont la connaissance est suffisante à déterminer sans équivoque les niveaux de motorisation Y . Par ailleurs, ces seuils n'apportent aucune information sur le niveau des variables latentes si leur vraisemblance n'est pas conditionnée par les choix Y . Ces deux remarques contribuent à simplifier la relation (3.44) en :

$$P(\Theta, S, Y^*, v | Y, X) \propto L(Y | Y^*, S) \times L(Y^* | \Theta, v, X) \times P(\Theta) \times P(S) \times P(v), \quad (3.45)$$

et font apparaître la vraisemblance $L(Y^* | \Theta, v, X)$ détaillée en (3.19). Toutes les distributions conditionnelles nécessaires à la mise en route de l'échantillonneur de Gibbs découlent de la relation précédente. Pour chaque paramètre (ou variable latente) considéré(e), ces distributions sont obtenues en réécrivant la relation (3.45) comme une simple fonction de celui-ci(ou celle-ci), les autres prenant la valeur de leur dernier tirage dans la boucle d'échantillonnage. En spécifiant des distributions *a priori* « conjuguées », nous obtenons les distributions conditionnelles suivantes. Nous explicitons dans un premier temps celles des paramètres, avant de nous intéresser à celles des variables individuelles. Nous adoptons les notations supplémentaires suivantes : soit π le vecteur colonne cumulé des coefficients du modèle dynamique, $\pi = (\beta', \gamma, \delta)'$, et $\Theta(-\theta)$, $S(-\theta)$ les ensembles de paramètres Θ et S privés du paramètre θ .

3.4.4.1 Distributions a posteriori des paramètres

- Distribution de $(\pi | \Theta(-\pi), S, Y^*, v, Y, X)$

La distribution d'une loi $\mathcal{N}(\underline{\pi}, \underline{V}_\pi)$ est utilisée comme *a priori* sur π . Afin d'alléger les écritures, nous utilisons la variable $Y_{it}^{*\pi} = Y_{it}^* - v_i$, et posons la matrice ligne $Z_{it} = (X_{it} \sim Y_{it-1}^* \sim (Y_{i1}^* - X_{i1}\beta_0))$ pour $t > 1$, où le signe \sim est ici l'opérateur de la concaténation horizontale. La distribution de $\pi | \Theta(-\pi), S, Y^*, v, Y, X$ est donnée par la relation (3.45) écrite comme une fonction de π : $P(\pi | \Theta(-\pi), S, Y^*, v, Y, X) \propto L(Y^* | \pi, \Theta(-\pi), v, X) \times P(\pi)$. Formellement, nous avons :

$$P(\pi \mid \Theta(-\pi), S, Y^*, v, Y, X) \propto \prod_i \prod_{t=2}^3 \exp\left(-\frac{1}{2\sigma_\varepsilon^2} (Y_{it}^{*\pi} - Z_{it}\pi)^2\right)^{w_i} \times \exp\left(-\frac{1}{2}(\pi - \underline{\pi})' \underline{V}_\pi^{-1} (\pi - \underline{\pi})\right).$$

Après quelques manipulations simples d'algèbre, nous obtenons cette relation de proportionnalité :

$$P(\pi \mid \Theta(-\pi), S, Y^*, v, Y, X) \propto \exp\left(-0.5(\pi - \bar{\pi})' \bar{V}_\pi^{-1} (\pi - \bar{\pi})\right),$$

avec :

$$\bar{V}_\pi = \left(\frac{1}{\sigma_\varepsilon^2} \sum_i \sum_{t=2}^3 w_i Z_{it}' Z_{it} + \underline{V}_\pi^{-1} \right)^{-1},$$

$$\bar{\pi} = \bar{V}_\pi \left(\frac{1}{\sigma_\varepsilon^2} \sum_i \sum_{t=2}^3 w_i Z_{it}' Z_{it} \times \hat{\pi} + \underline{V}_\pi^{-1} \times \underline{\pi} \right), \quad (3.46)$$

et où :

$$\hat{\pi} = \left(\sum_i \sum_{t=2}^3 w_i Z_{it}' Z_{it} \right)^{-1} \left(\sum_i \sum_{t=2}^3 w_i Z_{it}' Y_{it} \right).$$

Il en découle que la distribution conditionnelle *a posteriori* de π est normale :

$$\pi \mid \Theta(-\pi), S, Y^*, v, Y, X \sim \mathcal{N}(\bar{\pi}, \bar{V}_\pi). \quad (3.47)$$

Le résultat (3.46) possède un contenu intuitif très intéressant qui nous servira de raccourci prochainement. En se rappelant que l'inverse d'une variance est une mesure de la précision, on remarque notamment que l'espérance *a posteriori* n'est ici qu'une moyenne des estimateurs des moindres carrés et de l'espérance *a priori*, pondérée par leur précision. La variance *a posteriori* n'est autre que l'inverse de la somme des précisions des deux estimateurs, autrement dit, l'inverse de la précision totale.

- Distribution de $(\beta_0 \mid \Theta(-\beta_0), S, Y^*, v, Y, X)$

Nous utilisons une distribution $\mathcal{N}(\underline{\beta}_0, \underline{V}_{\beta_0})$ pour caractériser la connaissance *a priori* de β_0 , et définissons les variables $X_{i1}^{\beta_0} = -\delta X_{i1}$; $Y_{i2}^{*\beta_0} = Y_{i2}^* - X_{i2}\beta - (\gamma + \delta)Y_{i1}^* - v_i$; $Y_{i3}^{*\beta_0} = Y_{i3}^* - X_{i3}\beta - \gamma Y_{i2}^* - \delta Y_{i1}^* - v_i$ pour

alléger les prochaines écritures. La distribution de $\beta_0 \mid \Theta(-\beta_0), S, Y^*, v, Y, X$ découle de la relation (3.45) écrite comme une fonction de β_0 :

$$P(\beta_0 \mid \Theta(-\beta_0), S, Y^*, v, Y, X) \propto L(Y^* \mid \beta_0, \Theta(-\beta_0), v, X) \times P(\beta_0). \text{ Nous obtenons :}$$

$$\begin{aligned} P(\beta_0 \mid \Theta(-\beta_0), S, Y^*, v, Y, X) \propto \\ \prod_i \left(\prod_{t=2}^3 \exp \left(-\frac{1}{2\sigma_{\varepsilon_0}^2} (Y_{i1}^* - X_{i1}\beta_0)^2 \right) \times \exp \left(-\frac{1}{2\sigma_{\varepsilon}^2} (Y_{it}^{*\beta_0} - X_{i1}^{\beta_0}\beta_0)^2 \right) \right)^{w_i} \\ \times \exp \left(-\frac{1}{2} (\beta_0 - \underline{\beta}_0)' V_{\beta_0}^{-1} (\beta_0 - \underline{\beta}_0) \right). \end{aligned} \quad (3.48)$$

Puisque nous sommes dans une configuration similaire à celle de la distribution conditionnelle de π qui précède, c'est-à-dire une distribution *a priori* normale et conjuguée avec la vraisemblance, alors nous pouvons exploiter le raccourci déjà évoqué. De gauche à droite, l'expression (3.48) fait apparaître trois estimateurs de β_0 . Notamment, il résulte de la première exponentielle un premier estimateur $\hat{\beta}_0^a$ calculé par la formule des moindres carrés pondérés, et de précision $H^{\hat{\beta}_0^a}$:

$$\begin{aligned} \hat{\beta}_0^a &= \left(\sum_i w_i X_{i1}' X_{i1} \right)^{-1} \left(\sum_i w_i X_{i1}' Y_{i1}^* \right), \\ H^{\hat{\beta}_0^a} &= \sigma_{\varepsilon_0}^{-2} \left(\sum_i w_i X_{i1}' X_{i1} \right). \end{aligned}$$

De la même façon, il découle de la seconde exponentielle de (3.48) un autre estimateur de β_0 , noté $\hat{\beta}_0^b$ et de précision $H^{\hat{\beta}_0^b}$:

$$\begin{aligned} \hat{\beta}_0^b &= \left(\sum_i 2w_i X_{i1}^{\beta_0'} X_{i1}^{\beta_0} \right)^{-1} \left(\sum_i \sum_{t=2}^3 w_i X_{i1}^{\beta_0'} Y_{it}^{*\beta_0} \right), \\ H^{\hat{\beta}_0^b} &= 2\sigma_{\varepsilon}^{-2} \left(\sum_i w_i X_{i1}^{\beta_0'} X_{i1}^{\beta_0} \right). \end{aligned}$$

Enfin, l'estimateur de $\underline{\beta}_0$ résultant de la dernière exponentielle de la relation (3.48) est donné par l'espérance de la distribution *a priori* $\underline{\beta}_0$. Sa précision est calculée par l'inverse de la variance *a priori* : $V_{\beta_0}^{-1}$.

Compte tenu du produit d'exponentielles de la relation (3.48), la distribution *a posteriori* conditionnelle de β_0 est normale, d'espérance égale à la moyenne pondérée de ses estimateurs, et de variance égale à l'inverse de la somme des précisions :

$$\begin{aligned} (\beta_0 \mid \Theta_{-\beta_0}, S, Y^*, v, Y, X) &\sim \mathcal{N}(\bar{\beta}_0; \bar{V}_{\beta_0}), \\ \bar{V}_{\beta_0} &= \left(H^{\hat{\beta}_0^a} + H^{\hat{\beta}_0^b} + V_{\beta_0}^{-1} \right)^{-1}, \\ \bar{\beta}_0 &= \bar{V}_{\beta_0} \left(H^{\hat{\beta}_0^a} \hat{\beta}_0^a + H^{\hat{\beta}_0^b} \hat{\beta}_0^b + V_{\beta_0}^{-1} \beta_0 \right). \end{aligned} \quad (3.49)$$

- Distribution de $(\sigma_{\varepsilon_0}^{-2} \mid \Theta(-\sigma_{\varepsilon_0}^2), S, Y^*, v, Y, X)$

Il n'est pas inutile ici d'effectuer un bref rappel statistique. Une variable aléatoire Γ est distribuée selon une loi gamma de paramètres p et T , notée $\Gamma \sim \mathcal{G}(p, T)$ si sa fonction de densité G est de la forme :

$$G(x, p, T) \propto \exp(-Tx) x^{p-1},$$

pour p et T réels strictement positif. La variable aléatoire Γ est caractérisée par les moments suivants $E(\Gamma) = p/T$, $V(\Gamma) = p/T^2$.

Nous posons en *a priori* que $\sigma_{\varepsilon_0}^{-2} \sim \mathcal{G}(p_1/2, T_1/2)$. La densité de $\sigma_{\varepsilon_0}^{-2} \mid \Theta(-\sigma_{\varepsilon_0}^2), S, Y^*, v, Y, X$ découle de la relation (3.45) écrite comme une fonction de $\sigma_{\varepsilon_0}^{-2}$: $P(\sigma_{\varepsilon_0}^{-2} \mid \Theta(-\sigma_{\varepsilon_0}^2), S, Y^*, v, Y, X) \propto L(Y^* \mid \sigma_{\varepsilon_0}^{-2}, \Theta(-\sigma_{\varepsilon_0}^2), v, X) \times P(\sigma_{\varepsilon_0}^{-2})$.

En remplaçant par les formules, nous avons :

$$\begin{aligned} P(\sigma_{\varepsilon_0}^{-2} \mid \Theta(-\sigma_{\varepsilon_0}^2), S, Y^*, v, Y, X) &\propto \\ \prod_i (\sigma_{\varepsilon_0}^{-2})^{1/2} \exp\left(-\frac{\sigma_{\varepsilon_0}^{-2}}{2} (Y_{i1}^* - X_{i1}\beta_0)^2\right)^{w_i} &\times (\sigma_{\varepsilon_0}^{-2})^{(p_1/2)-1} \exp\left(-\frac{T_1}{2} \sigma_{\varepsilon_0}^{-2}\right). \end{aligned} \quad (3.50)$$

Moyennant un peu d'algèbre, le terme à droite dans la relation (3.50) égale :

$$\prod_i (\sigma_{\varepsilon_0}^{-2})^{\frac{W+p_1}{2}-1} \exp\left(-\frac{\sum_i w_i (Y_{i1}^* - X_{i1}\beta_0)^2 + T_1}{2} \sigma_{\varepsilon_0}^{-2}\right), \quad (3.51)$$

et par identification avec la fonction de densité de la loi gamma vue dans le rappel statistique précédent, nous concluons qu'*a posteriori*, la distribution conditionnelle de $\sigma_{\varepsilon_0}^{-2}$ est donnée par :

$$\sigma_{\varepsilon_0}^{-2} \mid \Theta(-\sigma_{\varepsilon_0}^2), S, Y^*, v, Y, X \sim \Gamma \left(\frac{W + p_1}{2}, \frac{\sum_i w_i (Y_{i1}^* - X_{i1} \beta_0)^2 + T_1}{2} \right). \quad (3.52)$$

- Distribution de $(\sigma_v^{-2} \mid \Theta(-\sigma_v^2), S, Y^*, v, Y, X)$

Toujours de la même façon, la distribution conditionnelle de σ_v^{-2} résulte de l'écriture de la relation (3.45) comme une fonction seule de σ_v^{-2} : $P(\sigma_v^{-2} \mid \Theta(-\sigma_v^2), S, Y^*, v, Y, X) \propto P(v) \times P(\sigma_v^{-2})$. En lui choisissant une distribution $\Gamma(p_2/2, T_2/2)$ comme *a priori*, on détaille ainsi cette la relation :

$$P(\sigma_v^{-2} \mid \Theta(-\sigma_v^2), S, Y^*, v, Y, X) \propto \prod_i (\sigma_v^{-2})^{1/2} \exp \left(-\frac{(v_i - 0)^2}{2} \sigma_v^{-2} \right)^{w_i} \times (\sigma_v^{-2})^{\frac{p_2}{2}-1} \exp \left(-\frac{T_2}{2} \sigma_v^{-2} \right), \quad (3.53)$$

et après simplification, le terme à droite dans la relation (3.53) devient :

$$(\sigma_v^{-2})^{\left(\frac{W+p_2}{2}-1\right)} \exp \left(-\frac{\sum_i w_i (v_i - 0)^2 + T_2}{2} \sigma_v^{-2} \right).$$

Il en découle par identification avec la densité d'une loi gamma que :

$$\sigma_v^{-2} \mid \Theta(-\sigma_v^2), S, Y^*, v, Y, X \sim \Gamma \left(\frac{W + p_2}{2}, \frac{\sum_i w_i (v_i - 0)^2 + T_2}{2} \right). \quad (3.54)$$

- Distribution de $(\sigma_{\varepsilon}^{-2} \mid \Theta(-\sigma_{\varepsilon}^2), S, Y^*, v, Y, X)$

Les conditions d'identification peuvent être interprétées comme des *a priori* infiniment puissants, de sorte qu'elles tiennent aussi lieu d'*a posteriori*. Autrement dit, il s'agit d'une certitude (qui ne peut donc être révisable *a posteriori*), traduite par une relation purement déterministe. Concernant $\sigma_{\varepsilon}^{-2}$, nous obtenons compte tenu de la condition (3.23) :

$$(\sigma_\varepsilon^{-2} \mid \Theta(-\sigma_\varepsilon^2), S, Y^*, v, Y, X) = (1 - \delta^2 \sigma_{\varepsilon_0}^2 - \sigma_v^2)^{-1}. \quad (3.55)$$

Concernant les *a posteriori* conditionnels des seuils $S = \{s_1, s_2, s'_1, s'_2\}$, nous examinons tout d'abord ceux du modèle général $\{s_1, s_2\}$, avant de considérer ceux du modèle initial $\{s'_1, s'_2\}$.

- Distribution de $(s_1 \mid \Theta, S(-s_1), Y^*, v, Y, X)$

Puisqu'il existe également une restriction d'identification portant sur s_1 en (3.23), alors celle-ci tient lieu d'*a posteriori* déterministe. Nous noterons toutefois qu'il s'agit d'une restriction qui ne dépend pas de la valeur d'autres paramètres :

$$(s_1 \mid \Theta, S(-s_1), Y^*, v, Y, X) = s_1 = 0. \quad (3.56)$$

- Distribution de $(s_2 \mid \Theta, S(-s_2), Y^*, v, Y, X)$

Il s'agit de la distribution conditionnelle la plus intuitive, mais peu évidente à dériver. Le point de départ incontournable est l'écriture de la relation (3.45), cette fois comme une fonction de s_2 , et qui donne : $P(s_2 \mid \Theta, S(-s_2), Y^*, v, Y, X) \propto L(Y \mid Y^*, s_2, S(-s_2)) \times P(s_2)$. Nous choisissons un *a priori* totalement non informatif pour s_2 en lui spécifiant une distribution uniforme, c'est-à-dire que $P(s_2) \propto c \times 1(s_1 < s_2 < +\infty)$. En détaillant, nous avons notamment :

$$\begin{aligned} L(Y \mid Y^*, s_2, S(-s_2)) &= \prod_{i,t} L(Y \mid Y^*, s_2, S(-s_2)) = \\ &= \prod_{i:Y_{i1}=0} 1(Y_{i1}^* \leq s'_1) \prod_{i:Y_{i1}=1} 1(s'_1 < Y_{it}^* \leq s'_2) \times \prod_{i:Y_{i1}=2} 1(Y_{it}^* > s'_2) \times \dots \\ &\dots \prod_{i,t \geq 2:Y_{it}=0} 1(Y_{it}^* \leq s_1) \prod_{i,t \geq 2:Y_{it}=1} 1(s_1 < Y_{it}^* \leq s_2) \times \prod_{i,t \geq 2:Y_{it}=2} 1(Y_{it}^* > s_2). \end{aligned}$$

En ne laissant paraître que des inégalités relatives à s_2 , nous pouvons écrire :

$$P(s_2 \mid \Theta, S(-s_2), Y^*, v, Y, X) \propto \prod_{i,t:Y_{it}=1} 1(Y_{it}^* \leq s_2) \times \prod_{i,t:Y_{it}=2} 1(Y_{it}^* > s_2) \times 1(s_1 < s_2 < +\infty).$$

Il en découle, de façon somme toute logique, que s_2 est distribué uniformément entre le maximum des variables latentes dans la population des $Y_{i,t \geq 2} = 1$ et le minimum des variables latentes dans la population des $Y_{i,t \geq 2} = 2$, puisque s_2 sert de limite entre ces deux choix discrets. En tenant compte de l'*a priori*, nous avons formellement :

$$s_2 \mid \Theta, S(-s_2), Y^*, v, Y, X \sim \mathcal{U} \left[\max \left(s_1; \max_{i,t \geq 2} (Y_{it}^* : Y_{it} = 1) \right); \min_{i,t \geq 2} (Y_{it}^* : Y_{it} = 2) \right]. \quad (3.57)$$

- Distributions de $(s'_l \mid \Theta, S(-s'_l), Y^*, v, Y, X)$

Les valeurs des seuils $s' = \{s'_1, s'_2\}$ de la période initiale sont dans notre modélisation soumises à des restrictions d'identification. Il découle notamment des restrictions d'identification (3.23) que :

$$\begin{aligned} (s'_1 \mid \Theta, S(-s'_1), Y^*, v, Y, X) &= s_1 = 0, \\ (s'_2 \mid \Theta, S(-s'_2), Y^*, v, Y, X) &= s_2. \end{aligned} \quad (3.58)$$

Après avoir considéré les *a posteriori* des paramètres et seuils, nous devons maintenant détailler ceux des variables propres à chaque ménage i , c'est-à-dire les facteurs d'hétérogénéité v_i et les variables latentes $Y_{i1}^*, Y_{i2}^*, Y_{i3}^*$.

3.4.4.2 Distributions a posteriori des variables individuelles

- Les variables latentes

On explicite les distributions conditionnelles des variables latentes Y_{it}^* de la même façon que celle des paramètres précédents, c'est-à-dire que l'on commence par écrire la relation (3.45) comme une fonction de Y_{it}^* . Pour tout i et $\forall t$ de 1 à 3, nous aboutissons à cette expression :

$$P(Y_{it}^* | \Theta, S, Y_{i-t}^*, v_i, Y_i, X_i) \propto L(Y_{it} | Y_{it}^*, S) L(Y_{it}^* | \Theta, Y_{i-t}^*, v_i, X_i), \quad (3.59)$$

où Y_{i-t}^* correspond au vecteur Y_i^* amputé de Y_{it}^* . Dans la relation (3.59), le résultat de $L(Y_{it} | Y_{it}^*, S)$ est une variable binaire qui vaut 1 si l'association $\{Y_{it}^*, S\}$ engendre effectivement la décision Y_{it} (0 sinon) en vertu de la règle de choix décrite par les systèmes (3.2) et (3.12). Formellement, il s'agit du résultat suivant :

$$L(Y_{it} | Y_{it}^*, S) = 1(a_{it} < Y_{it}^* < b_{it}), \quad (3.60)$$

avec a_{it} et b_{it} définis en (3.22), et où $1(\arg)$ désigne la fonction indicatrice valant 1 si l'argument est vrai, 0 sinon. Pour notre usage, $L(Y_{it} | Y_{it}^*, S)$ va seulement nous servir à poser les troncatures des distributions conditionnelles des variables latentes Y_{it}^* .

- Distribution de $(Y_{i1}^* | \Theta, S, Y_{i-1}^*, v_i, Y_i, X_i)$

L'écriture de (3.59) pour $t = 1$ donne lieu à la relation : $P(Y_{i1}^* | \Theta, S, Y_{i-1}^*, v_i, Y_i, X_i) \propto L(Y_{i1} | Y_{i1}^*, S) L(Y_{i1}^* | \Theta, Y_{i-1}^*, v_i, X_i)$. On procède à la définition des variables suivantes pour alléger les écritures ultérieures : $Y_{i3}^{*1} = Y_{i3}^* - X_{i3}\beta - \gamma Y_{i2}^* + \delta\beta_0 X_{i1} - v_i$, $Y_{i2}^{*1} = Y_{i2}^* - X_{i2}\beta + \delta\beta_0 X_{i1} - v_i$; puis nous pouvons écrire que :

$$\begin{aligned} L(Y_{i1}^* | \Theta, Y_{i-1}^*, v_i, X_i) &\propto \exp\left(-\frac{1}{2\sigma_{\varepsilon_0}^2} (X_{i1}\beta_0 - 1Y_{i1}^*)^2\right) \\ &\times \exp\left(-\frac{1}{2\sigma_{\varepsilon}^2} (Y_{i2}^{*1} - (\gamma + \delta)Y_{i1}^*)^2\right) \times \exp\left(-\frac{1}{2\sigma_{\varepsilon}^2} (Y_{i3}^{*1} - \delta Y_{i1}^*)^2\right). \end{aligned} \quad (3.61)$$

Compte tenu du produit d'exponentielles en (3.61), la distribution de $Y_{i1}^* | \Theta, Y_{i-1}^*, v_i, X_i$ est normale. Pour éviter les manipulations algébriques trop longues, nous pouvons adopter ici les mêmes mécanismes que ceux utilisés précédemment pour rendre explicite la distribution conditionnelle de β_0 . En appliquant l'estimateur des moindres carrés « pour chacune des exponentielles », nous calculons trois estimations de Y_{i1}^* : $\hat{Y}_{i1}^{*a} = X_{i1}\beta_0$, $\hat{Y}_{i1}^{*b} = (\gamma + \delta)^{-1} Y_{i2}^{*1}$,

$\hat{Y}_{i1}^{*c} = \delta^{-1} Y_{i3}^{*1}$; de précisions respectives : $H(\hat{Y}_{i1}^{*a}) = \sigma_{\varepsilon_0}^{-2}$, $H(\hat{Y}_{i1}^{*b}) = (\gamma + \delta)^2 \sigma_{\varepsilon}^{-2}$, $H(\hat{Y}_{i1}^{*c}) = \delta^2 \sigma_{\varepsilon}^{-2}$. En appliquant des méthodes maintenant familières, il vient que :

$$Y_{i1}^* \mid \Theta, Y_{i-1}^*, v_i, X_i \sim \mathcal{N}(\bar{Y}_{i1}^*, \bar{V}_{Y_1^*}), \quad (3.62)$$

avec :

$$\begin{aligned} \bar{V}_{Y_1^*} &= \left(H(\hat{Y}_{i1}^{*a}) + H(\hat{Y}_{i1}^{*b}) + H(\hat{Y}_{i1}^{*c}) \right)^{-1}, \\ \bar{Y}_{i1}^* &= \bar{V}_{Y_1^*} \left(H(\hat{Y}_{i1}^{*a}) \hat{Y}_{i1}^{*a} + H(\hat{Y}_{i1}^{*b}) \hat{Y}_{i1}^{*b} + H(\hat{Y}_{i1}^{*c}) \hat{Y}_{i1}^{*c} \right). \end{aligned}$$

De (3.60) et (3.62), il vient que :

$$Y_{i1}^* \mid \Theta, S, Y_{i-1}^*, v_i, Y_i, X_i \sim 1(a_{i1} < Y_{i1}^* < b_{i1}) \times \mathcal{N}(\bar{Y}_{i1}^*, \bar{V}_{Y_1^*}). \text{ Autrement dit :}$$

$$Y_{i1}^* \mid \Theta, S, Y_{i-1}^*, v_i, Y_i, X_i \sim \mathcal{NT}_{[a_{i1}, b_{i1}]}(\bar{Y}_{i1}^*, \bar{V}_{Y_1^*}). \quad (3.63)$$

- Distribution de $(Y_{i2}^* \mid \Theta, S, Y_{i-2}^*, v_i, Y, X)$

Pour $t = 2$, la distribution conditionnelle en (3.59) s'écrit $P(Y_{i2}^* \mid \Theta, s, Y_{i-}^*, v_i, Y_i, X_i) \propto L(Y_{i2} \mid Y_{i2}^*, s) L(Y_{i2}^* \mid \Theta, Y_{i-2}^*, v_i, X_i)$. Les formulations suivantes seront plus compactes si l'on définit ces variables : $Y_{i2}^{*2} = Z_{i2}\pi + v_i$, $Y_{i3}^{*2} = Y_{i3}^* - X_{i3}\beta - \delta(Y_{i1}^* - X_{i1}\beta_0) - v_i$. En détaillant le terme $L(Y_{i2}^* \mid \Theta, Y_{i-2}^*, v_i, X_i)$, nous avons :

$$L(Y_{i2}^* \mid \Theta, Y_{i-2}^*, v_i, X_i) \propto \exp\left(-\frac{1}{2\sigma_{\varepsilon}^2} (Y_{i2}^{*2} - Y_{i2}^*)^2\right) \times \exp\left(-\frac{1}{2\sigma_{\varepsilon}^2} (Y_{i3}^{*2} - \gamma Y_{i2}^*)^2\right). \quad (3.64)$$

Là aussi, la distribution conditionnelle de Y_{i2}^* en (3.64) est normale compte tenu du produit d'exponentielles. En appliquant les mêmes méthodes que précédemment, les deux exponentielles de (3.64) nous donnent deux estimations par les moindres carrés de Y_{i2}^* : $\hat{Y}_{i2}^{*a} = Y_{i2}^{*2}$ et $\hat{Y}_{i3}^{*b} = \gamma^{-1} Y_{i3}^{*2}$, de précisions respectives $\sigma_{\varepsilon}^{-2}$ et $\gamma^2 \sigma_{\varepsilon}^{-2}$. En se remémorant que dans ce cas normal, la variance est donnée par l'inverse de la précision totale des estimations par moindres carrés, et que l'espérance est donnée par la moyenne pondérée par leur précision de ces estimations, il en résulte que :

$$Y_{i2}^* \mid \Theta, Y_{i-2}^*, v_i, X_i \sim \mathcal{N}(\bar{Y}_{i2}^*, \bar{V}_{Y_{i2}^*}), \quad (3.65)$$

avec :

$$V_{Y_{i2}^*} = \left((1 + \gamma^2) \sigma_\varepsilon^{-2} \right)^{-1},$$

$$\bar{Y}_{i2}^* = V_{Y_{i2}^*} \left(\sigma_\varepsilon^{-2} Y_{i2}^{*2} + \gamma^2 \sigma_\varepsilon^{-2} (Y_{i3}^{*2} / \gamma) \right).$$

Sachant (3.60) et (3.65), nous obtenons la relation de proportionnalité $P(Y_{i2}^* | \Theta, S, Y_{i-2}^*, v_i, Y_i, X_i) \propto 1(a_{i2} < Y_{i2}^* < b_{i2}) \times \mathcal{N}(\bar{Y}_{i2}^*, \bar{V}_{Y_{i2}^*})$, qui nous autorise à conclure que la distribution conditionnelle de Y_{i2}^* est normale tronquée :

$$Y_{i2}^* | \Theta, s, Y_{i-2}^*, v_i, Y_i, X_i \sim \mathcal{NT}_{[a_{i2}, b_{i2}]}(\bar{Y}_{i2}^*, \bar{V}_{Y_{i2}^*}). \quad (3.66)$$

- Distribution de $(Y_{i3}^* | \Theta, S, Y_{i-3}^*, v_i, Y, X)$

La relation (3.59) écrite pour la dernière période $t = 3$ nous donne $P(Y_{i3}^* | \Theta, s, Y_{i-3}^*, v_i, Y_i, X_i) \propto L(Y_{i3}^* | Y_{i3}^*, s) L(Y_{i3}^* | \Theta, Y_{i-3}^*, v_i, X_i)$. Pour simplifier les écritures, nous posons : $\bar{Y}_{i3}^* = Z_{i3}\pi + v_i$, et en substituant, nous obtenons :

$$P(Y_{i3}^* | \Theta, S, Y_{i-3}^*, v_i, Y_i, X_i) \propto 1(a_{i3} < Y_{i3}^* < b_{i3}) \times \exp \left(-\frac{1}{2\sigma_\varepsilon^2} (Y_{i3}^* - \bar{Y}_{i3}^*)^2 \right).$$

Il en découle que le terme de droite est proportionnel à $1(a_{i3} < Y_{i3}^* < b_{i3}) \times \mathcal{N}(\bar{Y}_{i3}^*, \sigma_\varepsilon^2)$, et donc que la distribution conditionnelle de Y_{i3}^* est normale tronquée :

$$Y_{i3}^* | \Theta, S, Y_{i-3}^*, v_i, Y, X \sim \mathcal{NT}_{[a_{i3}, b_{i3}]}(\bar{Y}_{i3}^*, \sigma_\varepsilon^2). \quad (3.67)$$

- Les termes d'hétérogénéité

- Distribution de $(v_i | \Theta, S, Y_i^*, Y_i, X_i)$

L'écriture de la relation (3.45) comme seule fonction de v_i nous donne $P(v_i | \Theta, S, Y_i^*, Y_i, X_i) \propto L(Y_i^* | \Theta, v_i, X_i) \times P(v_i)$. En posant les variables : $Y_{it}^{*v} = Y_{it}^* - Z_{it}\pi$ pour $t \geq 2$, nous obtenons en remplaçant :

$$P(v_i | \Theta, S, Y_i^*, Y_i, X_i) \propto \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^3 (Y_{it}^{*v} - 1v_i)^2 \right) \times \exp \left(-\frac{1}{2\sigma_v^2} (v_i - 0)^2 \right).$$

Ce produit d'exponentielles nous assure que la variable $v_i \mid \Theta, s, Y_i^*, Y_i, X_i$ est distribuée selon une loi normale dont on calcule facilement la précision totale (la somme de la précision des deux estimations par moindres carrés) : $\bar{V}_v^{-1} = 2\sigma_\varepsilon^{-2} + \sigma_v^{-2}$; et l'espérance : $\bar{v}_i = \bar{V}_v^{-1} \left(2\sigma_\varepsilon^{-2} \left(\frac{1}{2} \sum_{t=2}^3 Y_{it}^{*v} \right) + \sigma_v^{-2} (0) \right)$. Formellement :

$$v_i \mid \Theta, S, Y_i^*, Y_i, X_i \sim N(\bar{v}_i, \bar{V}_v). \quad (3.68)$$

3.5 Sélection des variables et statistiques descriptives

Dans cette partie, nous présentons quelques statistiques descriptives du panel pondéré Parc Auto 1999-2001, qui font état de corrélations entre la motorisation des ménages et un ensemble de leurs caractéristiques. Le codage et le libellé des variables retenues pour notre étude, ainsi que l'information statistique synthétisée par des couples 'moyenne—écart-type' nous sont donnés dans la table 3.1.

La répartition des ménages entre les différents états de motorisation semble assez stable sur les trois années couvrant la période 1999-2001. Environ 2 ménages sur 10 n'ont pas de voiture, 5 ménages sur 10 en ont une, et 3 sur 10 sont multi-motorisés (table 3.2).

TABLE 3.1 : Variables d'intérêt, libellés, et statistiques descriptives

Variable	Libellé	Moyenne	(Ec-type)
nbvoi	nombre de voitures à la disposition du ménage	1.186	(0.813)
ageche	âge du chef de famille (CDF)	52.003	(16.483)
<i>Nombre dans le ménage de (CDF excepté)...</i>			
nbactoc	actifs occupés	0.394	(0.533)
nbretrai	retraités	0.130	(0.341)
nbfem	femmes adultes	0.983	(0.479)
<i>Nombre d'adultes âgés (CDF excepté)...</i>			
nbad1840	de 18 à 40 ans (hors enfants majeurs)	0.265	(0.443)
nage4160	de 41 à 60 ans	0.242	(0.428)
nage6170	de 61 à 70 ans	0.108	(0.313)
nage71p	de plus de 70 ans	0.068	(0.251)
<i>Nombre d'enfants âgés...</i>			
nage05	de moins de 6 ans	0.150	(0.441)
nage611	de 6 à 11 ans	0.171	(0.473)
nage1217	de 12 à 17 ans	0.181	(0.504)
enfmaj	de 18 à 30 ans	0.181	(0.505)
<i>Indicatrices : le chef de famille est un...</i>			
chefact	actif occupé par un emploi	0.593	(0.491)
chefret	retraité	0.350	(0.477)
chefchom	actif sans emploi	0.057	(0.232)
<i>Indicatrices : le revenu annuel net du ménage est...</i>			
r12	inférieur à 75kF	0.155	(0.362)
rev3	compris dans [75kF, 100kF[0.125	(0.331)
rev4	compris dans [100kF, 125kF[0.144	(0.351)
rev5	compris dans [125kF, 150kF[0.134	(0.340)
rev6	compris dans [150kF, 175kF[0.102	(0.303)
rev7	compris dans [175kF, 200kF[0.098	(0.297)
r89	compris dans [200kF, 300kF[0.174	(0.379)
r1013	supérieur ou égal à 300kF	0.067	(0.251)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.1 : Variables d'intérêt, libellés, et statistiques descriptives (suite)

Variable	Libellé	Moyenne	(Ec-type)
<i>Indicatrices: le ménage réside...</i>			
Paris	à Paris	0.050	(0.218)
PC	en petite couronne francilienne	0.060	(0.237)
GC	en grande couronne francilienne	0.068	(0.252)
llm	dans les agglomérations de Lille Lyon ou Marseille	0.056	(0.229)
ruraux	en Province, en zone rurale	0.031	(0.173)
periur	en Province, en zone périurbaine	0.330	(0.470)
banl	en Province, en banlieue	0.150	(0.358)
centre	en Province, en ville-centre	0.254	(0.435)
<i>Indicatrices : le ménage compte...</i>			
perm0	aucun permis de conduire	0.091	(0.288)
perm1	un permis de conduire	0.357	(0.479)
perm2	deux permis de conduire	0.473	(0.499)
perm3p	trois permis de conduire ou plus	0.079	(0.269)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.2 : Répartition annuelle des effectifs dans les états de motorisation

		Année		
Etat de motorisation (Y_{it})		1999	2000	2001
Non-motorisé ($Y_{it} = 0$)		570 (19.03%)	544 (18.16%)	554 (18.49%)
Mono-motorisé ($Y_{it} = 1$)		1508 (50.31%)	1510 (50.39%)	1487 (49.61%)
Multi-motorisé ($Y_{it} = 2$)		919 (30.66%)	943 (31.46%)	956 (31.91%)

Notes : pourcentages entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

Sur le panel empilé, on observe sans surprise un taux élevé de ménages non-motorisés lorsque le chef est au chômage (62%). A l'inverse, 87% des ménages dont le chef est un actif occupé sont motorisés, et parmi ceux-là, la proportion de ménages mono-motorisés est sensiblement inférieure à celle des multi-motorisés (respectivement 47% contre 53%). Si les ménages dont le chef est retraité sont également beaucoup plus souvent motorisés (78%) que ceux dont le chef est chômeur, ils se distinguent des ménages avec un chef actif occupé dans la répartition des motorisés : 77% d'entre eux n'ont qu'une seule voiture (table 3.3).

TABLE 3.3 : Répartition des ménages dans les états de motorisation selon l'activité du chef de famille

Etat de motorisation (Y_{it})		Statut d'activité du chef de famille		
		Sans emploi	Actif occupé	Retraité
Non-motorisé	($Y_{it} = 0$)	317.64 (61.94%)	676.16 (12.68%)	675 (21.45%)
Mono-motorisé	($Y_{it} = 1$)	145.60 (28.39%)	2452.01 (45.98%)	1907 (60.62%)
Multi-motorisé	($Y_{it} = 2$)	49.60 (9.67%)	2204.04 (41.33%)	564 (17.93%)

Notes : pourcentages entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.4 : Nombre d'automobiles dans le ménage selon le nombre de permis de conduire possédés

Nombres de permis	Effectifs de ménages	Moyenne	(Ec-type)
Aucun permis	818	0.101	(0.391)
Un permis	3205	0.802	(0.520)
Deux permis	4257	1.502	(0.626)
Trois permis (ou plus)	711	2.315	(0.839)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

Le nombre de permis de conduire dans le ménage est un déterminant fonctionnel patent de son niveau de motorisation, et nous constatons effectivement dans la table 3.4 une croissance des taux de motorisation d'environ 0.7 voiture par permis supplémentaire dans le ménage. Les statistiques descriptives de la table 3.5 nous montrent également sans surprise que les taux moyens de motorisation des ménages par zone sont inversement liés à la densité de la population : il y a globalement moins de voitures par ménage en Île-de-France qu'en Province. Dans le détail, les écarts de moyenne sont conséquents puisque les ménages provinciaux ruraux ont (en moyenne) 1.61 voitures à leur disposition contre 0.93 dans les métropoles de Lille, Lyon ou Marseille, 1.16 en grande couronne francilienne, et seulement 0.45 pour les ménages résidents à Paris.

TABLE 3.5 : Nombre d'automobiles par zone de résidence

Région	Zone de résidence	Effectifs de ménages	Moyenne	(Ec-type)
Province	Rural	279	1.612	(0.921)
	Périurbain	2969	1.473	(0.813)
	Banlieue	1352	1.280	(0.762)
	Ville-centre	2284	1.001	(0.710)
	Lille-Lyon-Marseille	501	0.926	(0.689)
Île-de-France	Grande couronne	613	1.158	(0.758)
	Petite couronne	541	0.840	(0.741)
	Paris	452	0.448	(0.650)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

Les taux moyens de motorisation par tranche d'âge du chef de famille croissent dans un premier temps pour atteindre un pic vers 41-50 ans (1.47 voitures par ménage). S'amorce ensuite une diminution plus rapide du nombre moyen de véhicules par ménage puisque ce taux moyen n'est plus que de 1.07 lorsque le chef du ménage est âgé d'entre 61 et 70 ans (table 3.6).

TABLE 3.6 : Nombre d'automobiles selon la tranche d'âge du chef de famille

Age du chef de famille (années)	Effectifs de ménages	Moyenne	(Ec-type)
18-30	826	1.091	(0.772)
31-40	1991	1.274	(0.736)
41-50	1592	1.472	(0.915)
51-60	1400	1.415	(0.860)
61-70	1712	1.073	(0.709)
71-80	1245	0.751	(0.624)
Plus de 80	225	0.565	(0.643)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

La taille du ménage est incontestablement un facteur déterminant de son niveau de motorisation. Un individu supplémentaire dans le ménage vient généralement augmenter sa demande de transport, notamment en automobile. C'est ainsi que les taux de motorisation croissent à mesure qu'augmente le nombre d'adultes dans le ménage : lorsque le ménage est mono-adulte, ce taux n'est que de 0.61 alors qu'il est de 2.06 pour les ménages de plus de 3 adultes (table 3.7). Ces taux sont plus élevés lorsque les adultes supplémentaires sont des actifs occupés (table 3.8). Quant à la présence d'enfants, elle semble favoriser la motorisation puisque le taux est de 1.06 lorsqu'il n'y a pas de mineur dans le ménage alors qu'il passe à 1.44 lorsqu'il y a un mineur dans le ménage (table 3.9).

TABLE 3.7 : Nombre d'automobiles par nombre d'adultes dans le ménage

Nombre d'adultes	Effectifs de ménages	Moyenne	(Ec-type)
Un	2721	0.612	(0.574)
Deux	5145	1.336	(0.681)
Trois	818	1.824	(0.874)
Quatre et plus	307	2.059	(1.210)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.8 : Nombre d'automobiles par nombre d'actifs occupés (hors chef de famille) dans le ménage

Nombre d'actifs (CDF excepté)	Effectifs de ménages	Moyenne	(Ec-type)
Aucun	5640	0.931	(0.702)
Un	3173	1.602	(0.778)
Deux	165	1.802	(1.184)
Trois et plus	13	2.492	(0.664)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.9 : Nombre d'automobiles par nombre d'enfants mineurs dans le ménage

Nombre d'enfants mineurs	Effectifs de ménages	Moyenne	(Ec-type)
Aucun	6461	1.064	(0.809)
Un	1072	1.441	(0.770)
Deux	1010	1.564	(0.715)
Trois et plus	448	1.490	(0.688)

Notes : écarts-types entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

Les caractéristiques économiques des ménages sont des déterminants incontournables de leur niveau de motorisation. Le revenu est bien sûr l'échelle la mieux adaptée à l'analyse de l'équipement automobile des ménages (HIVERT, 2001). Pour chacun, l'enquête Parc Auto renseigne une fourchette dans laquelle se situe leur revenu annuel net, et nous comptons notamment que 39% des ménages du panel dont le budget est inférieur à 100 kF/an ne sont pas motorisés, presque la moitié (49%) d'entre eux sont mono-motorisés, alors que seuls 12% sont multi-motorisés. A mesure que le revenu annuel croît, on observe une augmentation globale de la motorisation. Elle s'effectue tout d'abord en faveur de la mono-motorisation (55% des ménages gagnant de 100 à 200 kF/an), puis en faveur de la

multi-motorisation au détriment de la non et mono-motorisation (49% des ménages gagnant plus de 200 kF/an sont multi-équipés, table 3.10).

TABLE 3.10 : Répartition des effectifs dans les états de motorisation selon le revenu annuel net des ménages

Revenu annuel net	Niveau de motorisation	Effectifs de ménages	
Moins de 100kF	Non-motorisé	984	(38.98%)
	Mono-motorisé	1237	(48.97%)
	Multi-motorisé	304	(12.05%)
Compris entre [100kF ; 200kF[Non-motorisé	495	(11.52%)
	Mono-motorisé	2349	(54.67%)
	Multi-motorisé	1452	(33.81%)
200kF et plus	Non-motorisé	189	(8.72%)
	Mono-motorisé	920	(42.38%)
	Multi-motorisé	1061	(48.90%)

Notes : pourcentages entre parenthèses.

Source : panel pondéré 1999-2001 de Parc Auto.

Les transitions annuelles entre états suggèrent une dépendance temporelle : 93% des ménages non-équipés en 1999 le restent en 2000, et 95% d'entre eux en 2000 le restent en 2001. Considérant la mono et multi-motorisation, ces pourcentages sont du même ordre de grandeur (tables 3.11). Cette forme apparente d'immobilisme peut être le signe d'un comportement d'habitude ou d'ajustement dynamique, annonçant que l'état de motorisation individuel passé est un facteur explicatif du choix courant de motorisation. L'hypothèse semble réaliste puisque l'automobile n'est manifestement pas un bien flexible. En effet, l'ajustement quantitatif du parc des ménages induit par une variation de ses déterminants ne semble pas instantané, puisque l'acquisition ou la cession d'une automobile s'accompagne de démarches coûteuses en temps (réflexion, recherche...), comme énoncé en introduction.

TABLE 3.11 : Transitions des ménages entre états de motorisation 1999-2000 et 2000-2001

2000			
1999	Non-motorisé	Mono-motorisé	Multi-motorisé
Non-motorisé	92.9%	7.1%	0%
Mono-motorisé	0.9%	93%	6.1%
Multi-motorisé	0%	7.4%	92.6%

2001			
2000	Non-motorisé	Mono-motorisé	Multi-motorisé
Non-motorisé	95.3%	3.3%	1.4%
Mono-motorisé	2.3%	91.4%	6.3%
Multi-motorisé	0.9%	9.3%	90.5%

Notes : lecture en ligne (100%).

Source : panel pondéré 1999-2001 de Parc Auto.

Les corrélations entre variables explicatives créent des interférences pour déterminer l'importance réelle de chaque déterminant sur le choix de motorisation des ménages par des méthodes descriptives⁷⁰ et justifient notre recours à l'économétrie, l'objectif étant ici d'estimer conjointement l'effet propre de chaque facteur sur la décision de motorisation des ménages, puis de commenter les résultats avec une analyse « toutes choses égales par ailleurs ».

3.6 Résultats d'estimation

La mise en pratique de l'échantillonnage de Gibbs réclame de fixer les paramètres des distributions *a priori* de la section 3.4.4. Ils ont été choisis de façon telle que les *a priori* soient très diffus. Les valeurs suivantes $\underline{\pi} = \underline{\beta}_0 = 0$ ont notamment été utilisées : $\underline{V}_{\pi} = 100 \times I_{35}$; $\underline{V}_{\beta_0} = 100 \times I_{33}$; $p_1 = p_2 = 2$; $T_1 = T_2 = 0.5$.

⁷⁰ Par exemple, le revenu est corrélé à la CSP du chef de famille, au nombre d'adultes et d'actifs occupés dans le ménage... de sorte qu'il est difficile de séparer son influence réelle sur la motorisation de celui des autres variables corrélées.

Puis, l'algorithme d'échantillonnage doit être calé sur des valeurs d'initialisation. Nous avons estimé par maximum de vraisemblance les paramètres d'un modèle probit ordonné statique sur la vague d'enquête Parc Auto 1999, et avons utilisé les valeurs obtenues pour initialiser les paramètres β , β_0 et s_2 dans notre boucle d'échantillonnage de Gibbs, en $\beta^{(0)}$, $\beta_0^{(0)}$ et $s_2^{(0)}$ (table 3.12). Les paramètres dynamiques $\gamma^{(0)}$, $\delta^{(0)}$ ont été fixés à 0.1, alors que les termes des variances initialement choisis sont : $\sigma_{\varepsilon_0}^{2(0)} = 1$, $\sigma_v^{2(0)} = 0.1$. Les variables individuelles ont été initialisées en $v_i^{(0)} = 0$, $Y_{i1}^{(0)} = X_{i1}\beta_0^{(0)}$ et $Y_{it}^{(0)} = X_{it}\beta^{(0)} + \gamma^{(0)}Y_{it-1}^{(0)}$ pour $t > 1$.

TABLE 3.12 : Initialisation des paramètres β , β_0 , s_2

Variable	coefficient	Variable	coefficient	Variable	coefficient
constante	-4.767	nage71p	0.062	GC	1.257
ageche	0.085	nage05	0.210	llm	1.181
ageche²/100	-0.084	nage611	-0.090	ruraux	2.391
chefret	-0.004	nage1217	0.075	periur	1.865
chefchom	-0.521	rev3	0.420	banl	1.723
nbactoc	0.179	rev4	0.566	centre	1.424
nbretrai	0.067	rev5	0.650	perm1	1.742
nbfem	-0.177	rev6	0.727	perm2	2.845
enfinaj	0.098	rev7	0.833	perm3p	3.679
nbad1840	0.317	r89	0.923		
nage4160	0.067	r1013	1.021	s₂	2.547
nage6170	0.042	PC	0.683		

Notes : lexique des variables en table 3.1. Estimation du modèle probit ordonné statique.

Source : vague 1999 du panel pondéré Parc Auto 1999-2001.

Ainsi amorcé, notre algorithme a effectué 400000 cycles d'échantillonnage, dont nous avons éliminé les 100000 premiers qui correspondent à une phase de convergence. Parmi les 300000 derniers, nous avons conservé les paramètres avec une fréquence de un tous les dix cycles, puisque les valeurs successives échantillonnées présentaient une très forte corrélation en série. Au final, nous disposons pour chaque paramètre d'une chaîne de 30000 tirages, dont nous avons tout d'abord contrôlé les propriétés de convergence. Notamment, nous avons

appliqué le diagnostic de GEWEKE (1992), dont la statistique associée $|CD|$ est reportée. Pour chaque paramètre scalaire, celle-ci est inférieure à 1.96, suggérant que les chaînes ont individuellement toutes convergé. Nous pouvons donc considérer que nos tirages de paramètres sont effectivement issus de leur distribution *a posteriori*, et les exploiter.

La table 3.13 présente les résultats d'estimation du modèle dynamique (3.5). La première colonne montre les moyennes marginales empiriques *a posteriori*, la seconde présente les intervalles de crédibilité à 95%. La troisième donne les résultats de la statistique $|CD|$. A titre informatif, les mêmes statistiques pour le modèle initial (3.11) sont reportées dans la table 3.14.

Le modèle observe un pic de la motorisation pour les ménages dont le chef de famille est âgé de 37 ans. Toutes choses égales, notamment le revenu, le modèle montre que les ménages sont moins motorisés lorsque le chef est sans emploi que lorsqu'il est actif. En revanche, les estimations ne montrent pas de différence significative en comparant un ménage dont le chef est actif avec un autre dont le chef est retraité.

Concernant la structure du ménage autour de son chef, le modèle mesure que l'effet d'un adulte supplémentaire sur la motorisation est dégressif avec l'âge. Notamment, l'ajout d'un adulte âgé de 18 à 40 ans augmente significativement la motorisation du ménage, l'effet d'un adulte supplémentaire âgé de 41 à 60 ans est moindre, mais n'est plus significatif qu'au seuil de crédibilité à 90%. Au-delà de 61 ans, le modèle ne distingue plus d'effet notable. Que l'adulte supplémentaire soit actif, retraité ou sans emploi, ce détail n'influe pas significativement sur la motorisation du ménage.

Au contraire, le genre des adultes est une variable déterminante : relativement à un homme, l'effet du fait d'être une femme est une diminution significative de la motorisation du ménage. Par ailleurs, les résultats montrent qu'à toutes choses égales et quel que soit leur âge, le nombre d'enfants n'influence pas significativement le niveau de motorisation des ménages.

Le revenu annuel se révèle un déterminant décisif du niveau de motorisation. En comparaison avec un ménage doté d'un budget inférieur à 75 kF/an, les ménages plus riches sont significativement plus motorisés. Fort logiquement, le modèle constate un effet positif et croissant à mesure que le revenu augmente.

TABLE 3.13 : Résultats du modèle probit ordonné dynamique

Coefficients	Moyenne	Int. de crédibilité	CD
constante	-2.511	[-3.331 ; -1.708]	0.79
Variables d'âge du chef de ménage			
ageche	0.025	[0.001 ; 0.051]	0.81
ageche²/100	-0.034	[-0.058 ; -0.009]	0.78
Indicatrices du statut d'activité du chef de ménage (Référence : chefact)			
chefret	0.082	[-0.094 ; 0.259]	0.66
chefchom	-0.311	[-0.564 ; -0.065]	0.67
Structure et composantes du ménage			
nbactoc	0.099	[-0.023 ; 0.222]	0.54
nbretrai	-0.018	[-0.185 ; 0.146]	1.27
nbfem	-0.141	[-0.265 ; -0.018]	0.59
enfmaj	0.099	[-0.044 ; 0.240]	0.93
nbad1840	0.289	[0.089 ; 0.487]	0.96
nage4160	0.155	[-0.019 ; 0.331]	0.43
nage6170	0.124	[-0.077 ; 0.325]	0.82
nage71p	0.105	[-0.129 ; 0.338]	0.60
nage05	0.061	[-0.057 ; 0.180]	0.39
nage611	-0.033	[-0.132 ; 0.067]	0.06
nage1217	0.002	[-0.088 ; 0.091]	0.27
Indicatrices de la fourchette de revenu du ménage (Référence : r12)			
rev3	0.257	[0.076 ; 0.436]	0.46
rev4	0.388	[0.211 ; 0.566]	0.79
rev5	0.515	[0.330 ; 0.701]	0.86
rev6	0.637	[0.433 ; 0.845]	0.76
rev7	0.650	[0.448 ; 0.852]	0.86
r89	0.739	[0.540 ; 0.940]	0.84
r1013	0.942	[0.701 ; 1.187]	0.89

Notes : lexique des variables en table 3.1. Moyennes marginales empiriques *a posteriori*, intervalles de crédibilité à 95%, et statistique |CD| de Geweke.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.13 : Résultats du modèle probit ordonné dynamique (suite)

Coefficients	Moyenne	Int. de crédibilité	CD
Indicatrices de localisation du ménage (Référence : Paris)			
PC	0.477	[0.195 ; 0.760]	0.29
GC	1.085	[0.793 ; 1.383]	0.45
llm	0.914	[0.616 ; 1.221]	0.61
ruraux	1.970	[1.569 ; 2.388]	0.68
periur	1.501	[1.207 ; 1.809]	0.61
banl	1.285	[0.996 ; 1.581]	0.65
centre	1.132	[0.867 ; 1.406]	0.61
Indicatrices du nombre de permis dans le ménage (Référence : perm0)			
perm1	1.341	[1.088 ; 1.601]	0.77
perm2	2.126	[1.823 ; 2.435]	0.79
perm3p	3.082	[2.681 ; 3.498]	0.78
Coefficients dynamiques			
γ	0.516	[0.458 ; 0.570]	0.75
δ	0.229	[0.183 ; 0.275]	0.83
Seuils			
s_1	0	.	.
s_2	4.067	[3.920 ; 4.205]	0.77
Variances			
σ_v^2	0.234	[0.105 ; 0.411]	0.74
$\sigma_{\varepsilon 0}^2$	2.280	[2.012 ; 2.571]	0.75
σ_{ε}^2	0.646	[0.465 ; 0.787]	0.73

Notes : lexique des variables en table 3.1. Moyennes marginales empiriques *a posteriori*, intervalles de crédibilité à 95%, et statistique |CD| de Geweke.

Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.14 : Résultats du modèle probit ordonné de la période initiale

Coefficients	Moyenne	Inter. de crédibilité.	CD
constante	-7.417	[-8.688 ; -6.197]	0.98
Variables d'âge du chef de ménage			
ageche	0.135	[0.093 ; 0.178]	1.06
ageche ² /100	-0.134	[-0.175 ; -0.093]	0.96
Indicatrices du statut d'activité du chef de ménage (Référence : chefact)			
chefret	-0.004	[-0.305 ; 0.300]	0.12
chefchom	-0.810	[-1.248 ; -0.376]	1.23
Structure et composantes du ménage			
nbactoc	0.276	[0.067 ; 0.483]	1.79
nbretrai	0.105	[-0.190 ; 0.406]	0.18
nbfem	-0.285	[-0.505 ; -0.064]	0.24
enfmaj	0.156	[-0.105 ; 0.414]	1.57
nbad1840	0.506	[0.164 ; 0.850]	0.49
nage4160	0.117	[-0.184 ; 0.418]	1.40
nage6170	0.078	[-0.276 ; 0.437]	1.13
nage71p	0.105	[-0.328 ; 0.542]	0.40
nage05	0.325	[0.132 ; 0.519]	0.67
nage611	-0.142	[-0.315 ; 0.030]	0.52
nage1217	0.118	[-0.047 ; 0.284]	0.59
Indicatrices de la fourchette de revenu du ménage (Référence : r12)			
rev3	0.659	[0.364 ; 0.956]	0.43
rev4	0.892	[0.601 ; 1.186]	0.86
rev5	1.020	[0.724 ; 1.320]	0.72
rev6	1.142	[0.813 ; 1.466]	0.47
rev7	1.312	[0.975 ; 1.648]	0.52
r89	1.447	[1.141 ; 1.758]	0.79
r1013	1.600	[1.196 ; 2.018]	0.41
Indicatrices de localisation du ménage (Référence : Paris)			
PC	1.070	[0.575 ; 1.564]	0.77
GC	1.970	[1.494 ; 2.456]	0.80
llm	1.852	[1.364 ; 2.351]	0.77
ruraux	3.761	[3.142 ; 4.395]	0.81
periur	2.922	[2.494 ; 3.359]	0.73
banl	2.701	[2.258 ; 3.148]	0.83
centre	1.070	[0.575 ; 1.564]	0.77
Indicatrices du nombre de permis dans le ménage (Référence : perm0)			
perm1	2.684	[2.282 ; 3.099]	0.90
perm2	4.452	[4.022 ; 4.901]	0.96
perm3p	5.724	[5.120 ; 6.362]	0.92
Seuils*			
s ₁ '	0	.	.
s ₂ '	4.067	[3.920 ; 4.205]	0.77
Variance*			
$\sigma_{\varepsilon 0}^2$	2.280	[2.012 ; 2.571]	0.75

Notes : moyennes marginales empiriques *a posteriori*, intervalles de crédibilité à 95%, et statistique |CD| de Geweke. *Les valeurs des paramètres de seuils sont communes aux deux modèles, initial et dynamique. Puisqu'elle intervient dans le modèle dynamique, la variance $\sigma_{\varepsilon 0}^2$ de l'aléa initial est également reportée dans la table 3.13.

Source : panel pondéré Parc Auto 1999-2001.

Concernant les disparités géographiques, le modèle estime que les ménages ruraux sont significativement plus motorisés que les urbains résidents des villes-centres ou des banlieues. De façon plus générale, les résultats suggèrent que les niveaux de motorisation des ménages sont décroissants avec la densité de population de leur zone de résidence. C'est ainsi qu'en Île-de-France et toutes choses égales, un ménage parisien est statistiquement moins motorisé qu'un ménage francilien de petite couronne, à son tour moins motorisé qu'un ménage francilien de grande couronne. Cela traduit également qu'une diminution de la densité de circulation ainsi qu'une moins bonne couverture des transports collectifs rendent plus attractive, voire nécessaire, la possession automobile. Notamment en Province, les coefficients de la table 3.13 montrent qu'en moyenne, les résidents des métropoles « millionnaires » Lille, Lyon, et Marseille (dotées d'un métro notamment) sont les moins motorisés.

Mais surtout, le nombre de permis de conduire détenus par les ménages est le facteur le plus explicatif de leur motorisation. Plus particulièrement, le modèle estime que les passages d'aucun à un permis, d'un à deux permis, et de deux à trois permis s'accompagnent toujours d'une augmentation significative de leur motorisation. On remarque également une augmentation dégressive du coefficient avec le nombre de permis de conduire, pouvant suggérer que l'équipement d'un ménage augmente moins vite que son nombre de permis. Ce constat n'est pas totalement surprenant puisque l'automobile est un bien dont on peut partager l'usage à plusieurs permis au sein du ménage (PAPON et HIVERT, 2006).

Le coefficient de l'endogène retardée γ est positif et très significatif, traduisant un phénomène de dépendance temporelle de la motorisation. Explicitement, un ménage aura d'autant plus de chances d'être motorisé (resp. non-motorisé) à la période t qu'il était motorisé (resp. non-motorisé) en $t - 1$. La motorisation passée tend donc à retenir dans son état le choix de motorisation courant des ménages. Cet « effet mémoire » révèle la rigidité des ménages pour ajuster leur niveau d'équipement automobile à leurs déterminants courants, potentiellement due au recours à l'habitude pour le définir.

Enfin, la part de variance totale de l'erreur du modèle qui est imputable au terme d'hétérogénéité individuelle est estimée à $\hat{\sigma}_\alpha^2 / \sigma_\varepsilon^2 = (1 - \hat{\sigma}_\varepsilon^2) / 1 = 37\%$. Près d'un tiers du comportement non expliqué par les régresseurs utilisés tient donc, en France, à des spécificités individuelles plutôt qu'à des causes accidentelles.

Dans sa globalité, le modèle révèle de bonnes propriétés d'ajustement puisque la seule connaissance des variables explicatives X_i en 2001 lui permet de prévoir correctement 74,5% des observations la même année. Dans le détail, le modèle prédit respectivement 65%, 83% et 68% des cas de ménages non, mono et multi-motorisés en 2001⁷¹.

3.7 Exploitation des résultats

3.7.1 Temps d'ajustement des comportements de motorisation

Nous présentons tout d'abord quelques statistiques issues de la distribution marginale *a posteriori* du coefficient γ . La proportion en $t + s$ de l'effet de long terme latent induit par un changement permanent de caractéristiques à partir de t se calcule par $(1 - \gamma) \sum_{h=0}^s \gamma^h$. En moyenne, le modèle estime ainsi que la moitié (48.4%) de l'effet de long terme est pris en compte par les ménages dès le court terme dans l'ajustement de leur comportement d'équipement. Au bout d'un an, les trois quarts de cet effet de long terme sont déjà absorbés, et il est assimilé à plus de 95% par les ménages dans leur décision de motorisation après quatre années (table 3.15).

⁷¹ Calculs effectués à partir des probabilités individuelles de long terme lorsque les caractéristiques \underline{X}_i ont été fixées sur celles de 2001.

TABLE 3.15 : Proportion de l'effet de long terme d'un changement permanent considéré par les ménages

s	C	[95%]
0	48.4%	[43.0 ; 54.2]
1	73.4%	[67.5 ; 79.0]
2	86.3%	[81.5 ; 90.4]
3	92.9%	[89.4 ; 95.6]
4	96.3%	[94.0 ; 98.0]
5	98.1%	[96.6 ; 99.1]
6	99.0%	[98.0 ; 99.6]

Lecture : la part moyenne de l'effet à long terme d'un changement permanent de caractéristiques pris en considération par les ménages après s années est de C%. Intervalles de crédibilité à 95% entre crochets. Source : estimation du probit ordonné dynamique.

Par ailleurs, la statistique γ^s mesure le taux de survie après s périodes de l'effet initial latent d'un changement de caractéristiques accidentel⁷². En moyenne, le modèle estime que la survie de l'effet initial n'est plus que d'environ 25% deux ans après le changement, et qu'il aura disparu à 95% entre la quatrième et la cinquième année après celle du changement (table 3.16).

TABLE 3.16 : Taux de survie de l'effet initial d'un changement de caractéristiques occasionnel

s	S	[95%]
0	100%	.
1	51.6%	[45.8 ; 57.0]
2	26.6%	[21.0 ; 32.5]
3	13.7%	[9.6 ; 18.5]
4	7.1%	[4.4 ; 10.6]
5	3.7%	[2.0 ; 6.0]
9	0.3%	[0.1 ; 0.6]

Lecture : la part moyenne de l'effet d'un changement de caractéristiques accidentel encore pris en considération par les ménages après s années est de S%. Intervalles de crédibilité à 95% entre crochets. Source : estimation du probit ordonné dynamique.

⁷² C'est-à-dire non reproduit les périodes suivantes.

3.7.2 L'effet du changement résidentiel

Les propriétés dynamiques de notre modèle sont ensuite mises en avant par la simulation des effets de court et long terme d'un changement résidentiel des ménages. Pour cela, nous avons défini des ménages synthétiques calculés aux caractéristiques moyennes des zones géographiques, pour lesquels nous avons individuellement simulé un déménagement vers une zone francilienne (Paris, petite ou grande couronne). On considère que ces ménages moyens ont des caractéristiques stables dans le temps, si bien que la distribution de leur variable latente est à l'équilibre. Les probabilités de motorisation associées découlent donc du système (3.34) écrit pour $X_{i,\leq t} = \bar{X}_z$, où \bar{X}_z est le vecteur de caractéristiques du ménage synthétique de la zone z . Nous obtenons les systèmes de probabilités suivants (table 3.17).

TABLE 3.17 : Probabilités de motorisation des ménages synthétiques

Zone résidentielle	Etat de motorisation		
	non	mono	multi
Île-de-France			
Paris	0.64 [0.52 ; 0.74]	0.36 [0.26 ; 0.47]	0.00 [0.00 ; 0.00]
Petite couronne	0.24 [0.17 ; 0.32]	0.73 [0.66 ; 0.79]	0.03 [0.01 ; 0.05]
Grande couronne	0.05 [0.03 ; 0.08]	0.79 [0.74 ; 0.82]	0.16 [0.11 ; 0.22]
Province			
Lille, Lyon, Marseille	0.15 [0.10 ; 0.21]	0.79 [0.75 ; 0.83]	0.06 [0.03 ; 0.09]
Rurale	0.00 [0.00 ; 0.01]	0.48 [0.34 ; 0.61]	0.52 [0.38 ; 0.66]
Périurbain	0.01 [0.01 ; 0.01]	0.60 [0.56 ; 0.64]	0.39 [0.35 ; 0.43]
Banlieue	0.03 [0.02 ; 0.04]	0.74 [0.70 ; 0.78]	0.23 [0.19 ; 0.28]
Ville-centre	0.10 [0.08 ; 0.13]	0.81 [0.79 ; 0.83]	0.08 [0.06 ; 0.11]

Notes : ménages synthétiques évalués aux caractéristiques moyennes des zones résidentielles.

Intervalles de crédibilité à 95% entre crochets.

Source : panel pondéré Parc Auto 1999-2001.

En Île-de-France, le ménage synthétique parisien présente logiquement la plus forte probabilité d'être non-équipé (0.64), alors que les ménages moyens de petite et de grande couronne montrent une forte propension au mono-équipement. Chez ce dernier, le non-équipement devient presque nul, alors que la multi-motorisation apparaît statistiquement. En Province non urbaine, la non-motorisation est très peu répandue dans le comportement d'équipement des ménages synthétiques. Le ménage moyen des agglomérations de Lille, Lyon ou Marseille et celui des villes-centres révèlent les plus fortes propensions au non, et mono-équipement automobile. En Province et à mesure que l'on s'éloigne de ces secteurs urbains, en passant successivement par les zones de banlieue urbaine, périurbaine, puis rurale, la probabilité de mono-motorisation diminue au profit de la probabilité de multi-équipement, et lui devient même inférieure pour le ménage rural.

Sur la base des statistiques de la table 3.17, nous simulons pour les ménages synthétiques un changement permanent de résidence vers la région Île-de-France. Instantanément pour chacun, les multiplicateurs de court terme en (3.29) modifient la variable latente, et donnent lieu à un nouveau système de probabilités évalué via les formules en (3.35). Individuellement, le changement résidentiel permanent n'atteint son plein effet que dans le long terme, et fait intervenir ses multiplicateurs (3.32) pour déterminer la distribution stabilisée de leur variable latente. Sous l'hypothèse du gel des autres facteurs, on utilise les formules (3.37) qui calculent les probabilités de motorisation de long terme induit par le déménagement. Par différence de ses deux systèmes de probabilités avec le système initial renseigné dans la table 3.17, nous obtenons les effets marginaux de court et long terme (formules (3.36) et (3.38)). Ils sont reportés dans la table 3.18 pour les changements résidentiels intra Île-de-France, puis dans la table 3.19 pour les déménagements de la Province vers une zone francilienne.

Selon la table 3.18, le ménage synthétique parisien qui viendrait s'excentrer en Île-de-France vers ses couronnes ressent un très faible accroissement de sa multi-motorisation, dont la probabilité estimée ne « plafonne » qu'à 6% en grande couronne. Au contraire, sa propension à se mono-motoriser augmente franchement au détriment de la non-motorisation : elle croît en moyenne de +0.12 dans le court terme si sa destination est en petite couronne, et de +0.26 si elle est en grande couronne. A long terme, ces mesures respectives sont de +0.24 et +0.44. En observant le chemin inverse, le ménage de petite couronne renonce presque

sûrement à la multi-motorisation. Il diminue par ailleurs sa probabilité de mono-équipement de 0.09 en moyenne dans le court terme et 0.21 dans le long terme lorsqu'il s'installe à Paris. Lorsqu'il se « recentre », le ménage de grande couronne francilienne renonce rapidement à la multi-motorisation. A court terme, cela profite tout d'abord à sa mono-motorisation s'il déménage en petite couronne, puis exclusivement à sa non-motorisation dans le long terme. S'il déménage à Paris, alors il augmentera sa probabilité de non-équipement dès le court terme de 12 points (37 à long terme). D'une façon générale, l'hétérogénéité des ménages synthétiques franciliens montre que les effets marginaux ne sont pas symétriques lorsque l'on compare le déménagement d'un ménage z vers une zone z' , et celui d'un ménage z' vers une zone z .

TABLE 3.18 : Effets marginaux de court et long terme d'un déménagement intra Île-de-France

		Destination					
		Paris		PC		GC	
		CT	LT	CT	LT	CT	LT
Origine	Paris	non		-0.12 [-0.19 ; -0.05]	-0.25 [-0.38 ; -0.11]	-0.27 [-0.34 ; -0.20]	-0.50 [-0.61 ; -0.38]
		mono		0.12 [0.05 ; 0.18]	0.24 [0.10 ; 0.37]	0.26 [0.19 ; 0.34]	0.44 [0.32 ; 0.54]
		multi		0.00 [0.00 ; 0.00]	0.01 [0.00 ; 0.02]	0.01 [0.01 ; 0.02]	0.06 [0.04 ; 0.10]
	PC	non	0.10 [0.04 ; 0.17]	0.23 [0.10 ; 0.37]		-0.11 [-0.16 ; -0.06]	-0.18 [-0.26 ; -0.10]
		mono	-0.09 [-0.14 ; -0.04]	-0.21 [-0.33 ; -0.09]		0.07 [0.03 ; 0.12]	0.07 [0.02 ; 0.14]
		multi	-0.01 [-0.03 ; 0.00]	-0.02 [-0.04 ; -0.01]		0.04 [0.02 ; 0.05]	0.11 [0.06 ; 0.16]
	GC	non	0.12 [0.08 ; 0.17]	0.37 [0.26 ; 0.49]	0.05 [0.03 ; 0.08]	0.15 [0.09 ; 0.23]	
		mono	-0.01 [-0.06 ; 0.05]	-0.22 [-0.33 ; -0.11]	0.02 [0.00 ; 0.06]	-0.03 [-0.09 ; 0.03]	
		multi	-0.12 [-0.16 ; -0.08]	-0.16 [-0.21 ; -0.11]	-0.08 [-0.12 ; -0.04]	-0.13 [-0.19 ; -0.07]	

Notes : lexique des variables en table 3.1. Effets évalués à partir des probabilités d'équipement des ménages synthétiques. CT et LT pour court et long terme. Intervalles de crédibilité à 95% entre crochets. Source : panel pondéré 1999-2001 de Parc Auto.

TABLE 3.19 : Effets marginaux de court et long terme d'un déménagement vers l'Île-de-France

		Destination						
		Paris		PC		GC		
		CT	LT	CT	CT	LT	CT	
Origine	llm	non	0.18 [0.12 ; 0.24]	0.42 [0.29 ; 0.55]	0.07 [0.03 ; 0.11]	0.17 [0.08 ; 0.28]	-0.02 [-0.06 ; 0.01]	-0.05 [-0.12 ; 0.02]
		mono	-0.13 [-0.19 ; -0.08]	-0.37 [-0.48 ; -0.25]	-0.05 [-0.08 ; -0.02]	-0.13 [-0.22 ; -0.06]	0.01 [0.00 ; 0.04]	0.02 [-0.01 ; 0.05]
		multi	-0.04 [-0.07 ; -0.02]	-0.05 [-0.09 ; -0.03]	-0.03 [-0.05 ; -0.01]	-0.04 [-0.07 ; -0.02]	0.01 [-0.01 ; 0.03]	0.03 [-0.01 ; 0.08]
	ruraux	non	0.08 [0.04 ; 0.12]	0.48 [0.36 ; 0.60]	0.04 [0.02 ; 0.06]	0.24 [0.17 ; 0.33]	0.01 [0.01 ; 0.02]	0.06 [0.04 ; 0.10]
		mono	0.33 [0.20 ; 0.45]	0.03 [-0.15 ; 0.22]	0.30 [0.19 ; 0.40]	0.25 [0.10 ; 0.40]	0.20 [0.12 ; 0.29]	0.33 [0.19 ; 0.46]
		multi	-0.41 [-0.52 ; -0.29]	-0.51 [-0.65 ; -0.37]	-0.34 [-0.44 ; -0.24]	-0.49 [-0.63 ; -0.35]	-0.22 [-0.30 ; -0.13]	-0.39 [-0.54 ; -0.24]
	periur	non	0.07 [0.05 ; 0.10]	0.35 [0.24 ; 0.47]	0.04 [0.02 ; 0.05]	0.15 [0.10 ; 0.22]	0.01 [0.00 ; 0.01]	0.03 [0.01 ; 0.05]
		mono	0.21 [0.17 ; 0.25]	0.03 [-0.09 ; 0.13]	0.18 [0.14 ; 0.22]	0.19 [0.13 ; 0.24]	0.09 [0.05 ; 0.13]	0.16 [0.10 ; 0.22]
		multi	-0.29 [-0.33 ; -0.24]	-0.38 [-0.42 ; -0.34]	-0.22 [-0.26 ; -0.17]	-0.34 [-0.39 ; -0.30]	-0.10 [-0.14 ; -0.06]	-0.19 [-0.26 ; -0.11]
	banl	non	0.12 [0.08 ; 0.16]	0.40 [0.29 ; 0.52]	0.06 [0.04 ; 0.08]	0.18 [0.12 ; 0.25]	0.01 [0.00 ; 0.02]	0.02 [0.00 ; 0.05]
		mono	0.06 [0.01 ; 0.10]	-0.18 [-0.30 ; -0.07]	0.07 [0.04 ; 0.10]	0.01 [-0.05 ; 0.07]	0.03 [0.00 ; 0.06]	0.05 [0.00 ; 0.09]
		multi	-0.17 [-0.21 ; -0.13]	-0.22 [-0.27 ; -0.18]	-0.13 [-0.16 ; -0.09]	-0.19 [-0.24 ; -0.15]	-0.04 [-0.07 ; 0.00]	-0.07 [-0.14 ; 0.00]
centre	non	0.20 [0.14 ; 0.26]	0.50 [0.38 ; 0.61]	0.10 [0.07 ; 0.13]	0.25 [0.16 ; 0.34]	0.01 [-0.02 ; 0.03]	0.01 [-0.03 ; 0.06]	
	mono	-0.13 [-0.19 ; -0.08]	-0.41 [-0.53 ; -0.30]	-0.05 [-0.08 ; -0.03]	-0.18 [-0.26 ; -0.10]	0.00 [-0.01 ; 0.00]	0.00 [-0.02 ; 0.00]	
	multi	-0.07 [-0.08 ; -0.05]	-0.08 [-0.10 ; -0.06]	-0.05 [-0.07 ; -0.03]	-0.07 [-0.09 ; -0.05]	0.00 [-0.02 ; 0.01]	-0.01 [-0.04 ; 0.03]	

Notes : lexique des variables en table 3.1. Effets évalués à partir des probabilités d'équipement des ménages synthétiques. CT et LT pour court et long terme. Intervalles de crédibilité à 95% entre crochets. Source : panel pondéré 1999-2001 de Parc Auto.

De la table 3.19 découlent les constatations suivantes. Quelque soit l'horizon temporel, le ménage synthétique des métropoles Lille, Lyon et Marseille conserve sensiblement le même comportement de motorisation lorsque l'on simule son déménagement en grande couronne francilienne. Par ailleurs, on remarque que le ménage rural, pourtant le plus enclin à la multi-motorisation dans sa zone d'origine, consent à abandonner dès le court terme une très forte propension à être multi-équipé lorsqu'il vient résider en Île-de-France. Dans les couronnes franciliennes, il tend à devenir essentiellement mono-motorisé, et le même constat vaut pour le ménage synthétique périurbain. Ce dernier montre des effets marginaux à court et long terme très voisins lorsqu'il s'installe en petite couronne. Lorsqu'il emménage en grande couronne, le ménage moyen de la zone banlieue tend à favoriser le mono-équipement et ne recourt pas à la non-motorisation. Exceptés le ménage moyen des villes-centres et celui des métropoles Lille-Lyon-Marseille, les autres ménages provinciaux dont on a simulé le déplacement à Paris augmentent dans le court terme leur probabilité d'être mono-équipé au détriment du multi-équipement, puis la diminue sur le long terme au profit du non-équipement automobile. Dans ces cas de figure, l'état de mono-équipement tient lieu d'étape intermédiaire entre multi-équipement et déséquipement.

3.7.3 Les élasticités au revenu

Comme dans HANLY et DARGAY (2000), nous souhaitons mettre en lumière les effets de revenu sur le niveau d'équipement des ménages, tout en tirant profit de notre spécification dynamique, en calculant les élasticités de court et long terme. À partir de WINSTON (1981) notamment, les élasticités au revenu de l'équipement automobile sont définies par les variations relatives des probabilités moyennes de choisir chaque état de motorisation, rapportées au taux de croissance du revenu. Avec nos notations, nous définissons donc les élasticités au revenu de court et long terme, $\varsigma_R^{CT}(j)$ et $\varsigma_R^{LT}(j)$, suivantes :

$$\varsigma_R^{CT}(j) = \frac{\frac{1}{\bar{W}} \sum_i w_i \Lambda_{ij}^{CT}(rR_{it})}{\frac{1}{\bar{W}} \sum_i w_i \Pr[Y_{it} = j \mid X_{i,\leq t}]} \times \frac{1}{r}, \quad (3.69)$$

$$\varsigma_R^{LT}(j) = \frac{\frac{1}{W} \sum_i w_i \Lambda_{ij}^{LT}(rR_{it})}{\frac{1}{W} \sum_i w_i \Pr[Y_{it} = j | X_{i,\leq t}]} \times \frac{1}{r}, \quad (3.70)$$

avec r le taux de croissance de la variable continue de revenu R . Malheureusement, notre information sur les revenus des ménages n'est constituée que d'indicatrices de tranches, et non de R directement. Pour calculer les élasticités (3.69) et (3.70), nous devons donc reconstituer les revenus. Pour cet exercice, nous utilisons une méthode d'imputation simple, en tirant individuellement les revenus dans une loi uniforme, dont l'intervalle de définition correspond aux bornes de la tranche de revenu déclarée⁷³. Les revenus simulés sont ensuite augmentés de 20% ($r = 0.2$), puis sont à nouveau discrétisés dans les tranches de départ. Les statistiques (3.69) et (3.70) sont calculées aux caractéristiques des ménages en 2001, et sont reportées dans la table 3.20.

TABLE 3.20 : Elasticités au revenu des probabilités d'équipement automobile

	Non-équipement	Mono-équipement	Multi-équipement
Court terme	-0.19 [-0.23 ; -0.15]	-0.05 [-0.07 ; -0.02]	0.18 [0.13 ; 0.23]
Long terme	-0.38 [-0.45 ; -0.31]	-0.11 [-0.16 ; -0.05]	0.37 [0.27 ; 0.46]

Notes : calculs des formules (3.69) et (3.70) aux caractéristiques X des ménages en 2001.

Intervalles de crédibilité à 95% entre crochets.

Sans surprise, nos résultats d'estimation font apparaître que la croissance du revenu des ménages profite à l'augmentation de leur motorisation automobile. A court ou à long terme, il en découle que l'élasticité au revenu de la probabilité de non-équipement est négative, et réciproquement, que celle de l'équipement est positive. L'automobile appartient donc, en France, à la catégorie des biens microéconomiques dits « normaux ». Plus précisément, les sensibilités à court

⁷³ Bien évidemment, il existe d'autres méthodes d'imputation beaucoup plus sophistiquées et précises. Par exemple, on aurait pu « reconstruire » le revenu des ménages à l'aide des résultats d'un modèle – probit ou logit – ordonné et de résidus simulés tronqués.

terme des probabilités de non et multi-motorisation sont de signes opposés comme attendu, et d’amplitude très similaires (respectivement de -0.19 et +0.18, table 3.20). Que celle du mono-équipement soit peu affectée par le revenu à court terme n’est pas surprenant puisqu’il s’agit de la modalité d’équipement la plus répandue dans la population. A long terme, les effets au revenu sont plus élevés et correspondent assez précisément au double des élasticités de court terme. Ce rapport tient largement au fait que le coefficient γ est proche de 0.5, et en conséquence, que les multiplicateurs latents de long terme sont globalement deux fois plus élevés que ceux de court terme (section 3.3.1).

3.8 Conclusion

Ce chapitre a traité du comportement quantitatif d’équipement automobile des ménages en France. Dans un environnement statistique bayésien, le modèle probit ordonné a été ajusté sur les observations du panel pondéré Parc Auto 1999-2001, pour caractériser trois états : la non-motorisation, la mono-motorisation, et la multi-motorisation. Grâce à la méthode d’estimation MCMC par échantillonnages de Gibbs, qui reconstitue notamment les variables latentes par augmentation de données, il nous a été possible de faire reposer notre modèle sur une forme latente autorégressive. Pour en approximer les conditions initiales, nous nous sommes inspirés de la méthode de l’estimateur linéaire dynamique de BLUNDELL et SMITH (1991). Originale dans ce contexte, cette forme autorégressive du modèle linéaire latent se démarque des études précédentes qui, au mieux, introduisaient la dynamique dans leur modèle latent par des indicatrices du niveau d’équipement passé. En comparaison, notre spécification permet d’adapter quelques outils d’analyse des séries temporelles à notre problème de choix catégoriel du niveau de motorisation, élargissant ainsi l’éventail des conclusions dynamiques.

Si, comme les études précédentes, notre modèle conclut également à la dynamique des comportements d’équipement automobile, l’inférence statistique est enrichie par l’analyse de court et long terme. Par exemple, notre modèle mesure que les ménages confrontés à un changement permanent de caractéristiques considèrent dès le court terme environ la moitié de l’effet total de long terme. Par ailleurs, le calcul des effets marginaux simulant un déménagement des ménages moyens franciliens vers une autre zone francilienne montre que le ménage parisien augmente sa probabilité d’équipement (initialement de 36%) de 27 points à court terme, puis

de 50 points à long terme lorsqu'il s'implante dans la grande couronne d'Île-de-France. Inversement, le ménage moyen de grande couronne francilienne qui s'installe à Paris diminue en moyenne sa probabilité d'équipement de 12 points à court terme, et 37 points à long terme. D'une façon générale et quel que soit l'horizon, les effets marginaux dans cette étude ne sont pas symétriques lorsque l'on observe deux ménages franciliens moyens, l'un déménageant dans la zone de l'autre et inversement, la raison tenant ici à leur hétérogénéité observée. Enfin, les résultats ne décèlent aucune modification significative à court ou long terme du comportement d'équipement lorsque le ménage moyen des agglomérations de Lille-Lyon-Marseille et celui des autres villes-centres de Province s'établissent en grande couronne francilienne.

La reconstitution d'un revenu annuel continu des ménages puis la simulation de sa croissance a permis de dériver les élasticités au revenu à court et long terme des probabilités d'équipement. Il est apparu que l'automobile est un bien normal, et que les probabilités d'équipement extrêmes (celles de non et multi-équipement) avaient à court terme des sensibilités au revenu opposées et d'amplitude équivalente. Ce constat tient également pour les sensibilités de long terme, qui sont globalement deux fois plus fortes que celles de court terme.

Outre le revenu et la zone de résidence, les résultats de la modélisation mettent également en évidence le rôle prédominant du nombre de permis de conduire dans le ménage pour expliquer son niveau d'équipement.

La qualité de l'équipement des ménages n'est pas traitée dans cette étude. Pourtant, on pourrait, en parallèle avec une saturation des besoins en automobiles, soupçonner son effet lorsqu'on observe la croissance globalement concave des coefficients des tranches de revenu (table 3.13). Associé aux élasticités-revenu inférieures à un⁷⁴ (table 3.20), ceci suggère que le niveau de motorisation augmente moins vite que le revenu, et que les ménages le compensent par exemple en améliorant la qualité de leur parc (avec des voitures moins âgées, moins énergivores, plus confortables, plus puissantes...) Intuitivement, il n'y a aucune difficulté à présumer que les augmentations de revenu des ménages s'accompagnent d'une augmentation qualitative de leurs voitures lorsque le nombre d'automobiles devient égal à celui de permis de conduire. Un prolongement nécessaire pour comprendre le comportement d'équipement automobile des ménages réside donc dans la prise en compte que la voiture est un bien très différencié (plusieurs marchés, plusieurs

⁷⁴ En valeur absolue.

niveaux de gamme, plusieurs carburants...). Dans le prochain chapitre, nous aborderons cet angle qualitatif du parc automobile en France, en étudiant les déterminants du choix des agents lorsqu'ils font face à plusieurs modalités discrètes d'acquisition.

Chapitre 4 :

Etude du choix d'acquisition automobile

4.1 Introduction

Sur la scène internationale, la France a le privilège d'appartenir au club des pays producteurs d'automobiles. Forte de cet avantage, l'industrie automobile constitue un fleuron de son économie, dans laquelle les constructeurs sont des acteurs majeurs qui créent de la richesse et de l'emploi. Le secteur automobile est caractérisé par une sensibilité à la conjoncture économique particulièrement forte, si bien que son état de santé est particulièrement surveillé par les pouvoirs publics. Quand celui-ci entre en crise, des mesures politiques visant à revigorer l'acquisition d'automobiles neuves pour soutenir la production et l'emploi peuvent être considérées par le législateur. Deux exemples célèbres de soutien en France sont les '*balladurettes*' (1994-1995) et '*juppettes*' (1995-1996), qui subventionnaient l'achat de voitures neuves contre la mise au rebut des voitures les plus âgées (ADDA et COOPER, 2000). En incitant ainsi le renouvellement, ces mesures prenaient également une dimension écologique, en évinçant les « mauvais élèves » du parc automobile français, les vieilles voitures étant plus gourmandes en carburant et plus polluantes. Côté demande, l'automobile neuve reste un bien durable relativement cher qui, plus qu'un achat, nécessite un « investissement » par les agents. Ainsi, l'intérêt individuel pour l'offre de ces voitures neuves est clairement hétérogène entre les agents au moment de l'acquisition, certains privilégiant alternativement de s'équiper en ayant recours au marché de l'occasion, qui vient rallonger la durée de vie des automobiles.

Les préoccupations écologiques ne portent pas uniquement sur l'âge ou sur le choix du marché pour l'acquisition des voitures : selon qu'elles fonctionnent aux supercarburants ou au gazole, la contribution des gaz d'échappement à la dégradation de la qualité de l'air est différente. Au moins pour les automobiles récentes et à niveau de gamme équivalent, les voitures diesel sont moins émettrices de gaz à effet de serre au kilomètre, alors qu'elles rejettent davantage de particules que les voitures essence. De plus, le bilan écologique se détériore à mesure que le niveau de gamme augmente, puisque celui-ci est positivement corrélé avec la taille de l'automobile et de sa cylindrée.

En addition à ces considérations d'ordre écologique, la typologie des véhicules tient un rôle en matière de sécurité, les véhicules de haut de gamme protégeant vraisemblablement davantage les automobilistes en cas d'accident que les voitures plus modestes. Elle tient également un rôle dans la gestion de l'espace

occupé par l'automobile dans les zones fortement urbanisées, particulièrement en Île-de-France où les enjeux de la fluidité du trafic routier et du stationnement sont récurrents, lorsque les « petites urbaines » se révèlent bien plus manipulables et moins encombrantes dans ce contexte que les « grandes routières ».

Au cœur de ces préoccupations se pose donc la question de la qualité du parc automobile, mais aussi des usagers qui le composent, en choisissant leurs voitures.

Les véhicules récents, moins bruyants, moins gourmands en carburant, mais aussi plus sûrs et plus écologiques que les anciens modèles, marquent un réel effort des constructeurs visant à limiter les effets négatifs de l'usage de la voiture sur le budget et le confort de vie des automobilistes, mais aussi de la collectivité en général.

Les politiques publiques qui visent à réduire les nuisances automobiles en accélérant le renouvellement du parc en faveur des véhicules les plus écologiques ont besoin d'une meilleure connaissance des comportements qualitatifs d'équipement. Par ailleurs, révéler les inégalités d'accès et de préférences à l'ensemble des véhicules offerts sur les marchés (ou construire une typologie des automobilistes) est un exercice pour lequel les constructeurs présentent un intérêt continu, et qui nécessite l'analyse du choix individuel et qualitatif d'acquisition automobile.

Sommairement, ce quatrième chapitre s'attache à répondre à la question triviale suivante : « qui choisit quoi ? ». Plus spécifiquement, nous mesurons le poids des caractéristiques individuelles, telles que la catégorie professionnelle (en proxy du revenu), le sexe, et l'âge dans la décision des agents au moment de choisir une automobile parmi un ensemble. Nous tenterons de mettre en lumière des singularités d'équipement selon la zone de résidence des agents, en confrontant notamment les résidents franciliens et les Provinciaux. Le découpage géographique en couronnes de la région Île-de-France permettra d'évaluer plus finement les différences comportementales d'acquisition entre Franciliens. Rappelons que le second chapitre de la thèse traitait des répercussions de la variation du prix des carburants sur l'automobilité des ménages : des résultats viendront compléter cette analyse en examinant ces effets sur le choix des agents à l'acquisition d'une automobile. Par ailleurs, l'habitude est un facteur explicatif souvent pertinent pour décrire la pérennité des choix dans le temps. Dans les pages suivantes, nous examinons la fidélité des agents aux caractéristiques physiques de leur ancien véhicule (lorsqu'il existe) pour valider l'hypothèse de dépendance temporelle des

choix d'automobiles, permettant de conclure sur la dynamique des comportements d'acquisition des agents.

Sur la base d'observations désagrégées au niveau individuel, un modèle économétrique approprié pour décrire ce comportement d'acquisition doit appartenir à la famille des modèles de choix discrets. Dans ce chapitre, nous visitons plus précisément le modèle probit multinomial (MNP⁷⁵).

Par le passé, ce modèle a été écarté des méthodes d'analyse des choix qualitatifs automobiles, les auteurs lui préférant les déclinaisons de modèles logit, plus simples à formuler et à mettre en œuvre. Avec l'accroissement des capacités informatiques et les innovations statistiques, ces modèles logit se sont progressivement complexifiés, alors que le modèle MNP devenait plus accessible. Pourtant, les chercheurs ayant investi le champ d'analyse des choix d'automobiles sont restés très fidèles aux modèles logit, comme en témoigne notre revue de la littérature. Celle-ci est présentée dans une première section. Dans la suivante, nous exposons quelques fondements du choix discret relevant de la théorie microéconomique classique du consommateur. La description du modèle MNP, puis sa version bayésienne sont ensuite abordées dans une troisième section. La partie qui suit (4^{ème} section) traite de la résolution du modèle bayésien par une méthode « MCMC » (Monte Carlo Markov Chain) hybride décrite dans NOBILE (1998), qui combine les algorithmes d'échantillonnage de Gibbs et de Metropolis-Hastings. Les diagnostics de convergence de GELMAN et RUBIN (1992), puis de BROOKS et GELMAN (1998) sont présentés dans une cinquième partie. Nous appliquons ces outils sur le cumul des occurrences d'acquisition issues des vagues annuelles de 1997 à 2001 de l'enquête Parc Auto (6^{ème} section). Les possibilités d'acquisition automobile ont été rangées en huit modalités, croisant les caractéristiques d'état à l'achat, de carburant utilisé, et de gamme. Les résultats de l'ajustement du modèle MNP bayésien sont ensuite décrits (7^{ème} section), puis exploités dans une dernière partie par une analyse « *ceteris paribus* ».

4.2 Revue de littérature

Depuis les années 1980, plusieurs auteurs ont penché sous l'angle qualitatif leurs expertises du choix automobile. LAVE et TRAIN (1979) décryptent l'achat de

⁷⁵ Sigle anglosaxon pour « MultiNomial Probit model ».

voitures aux Etats-Unis, en observant le choix d'acquisition d'environ 500 ménages urbains parmi un ensemble de 10 modalités. Pour l'expliquer, les auteurs appliquent un modèle logit multinomial sur des combinaisons de variables faisant intervenir les caractéristiques des ménages et les attributs des modalités automobiles. SHERMAN et MANSKI (1980) utilisent également ce modèle pour décrire la possession automobile de 1200 ménages américains équipés en 1976, selon les critères de la marque, du modèle, et de l'année. L'espace des choix individuels consiste en l'automobile possédée, plus 25 automobiles sélectionnées au hasard parmi 600 véhicules. MANNERING et WINSTON (1985) proposent une approche globale des comportements automobiles américains, qui analyse le nombre, le type, et l'usage dans un cadre unifié. Concernant plus particulièrement la possession automobile, les auteurs appliquent un modèle logit sur un échantillon de 4000 ménages mono et bi-motorisés. Leurs possibilités de choix consistent en un ensemble de 10 véhicules caractérisés par la marque, le modèle et le millésime de fabrication : la modalité choisie augmentée de neuf automobiles issues d'un tirage aléatoire parmi 2000 véhicules. BROWNSTONE et al. (1996) utilisent les préférences déclarées des consommateurs et produisent les résultats d'un modèle logit multinomial qui permettent d'éclairer le pouvoir de pénétration des véhicules utilisant des carburants alternatifs sur le marché automobile. KITAMURA et al. (2000) appliquent le modèle logit multinomial sur un échantillon d'environ 2000 observations de ménages motorisés, recueilli dans l'aire urbaine de Los Angeles en 1993. Les auteurs s'attachent principalement à mettre en lumière l'importance du contexte résidentiel, des caractéristiques de l'usager principal du véhicule, et de celles de son ménage sur le type d'automobile possédée, lorsque l'espace des choix compte six classes de véhicules. CHOO et MOKTARIAN (2000) exploitent les retours d'environ 2000 questionnaires électroniques d'individus motorisés dans la baie de San Francisco. Pour expliquer le type d'automobile possédée, les auteurs utilisent des variables subjectives qui décrivent notamment la perception des personnes interrogées à propos de leur mobilité, de leur personnalité, et de leur mode de vie. Après avoir testé plusieurs schémas d'emboîtement à partir de neuf classes de véhicules, le modèle finalement retenu dans leur rapport est un logit multinomial. MOHAMMADIAN et MILLER (2003) analysent le comportement d'achat de véhicules des ménages canadiens entre 1990 et 1998. A partir d'un échantillon comptant environ 600 observations, les auteurs appliquent un modèle logit emboîté ayant cette structure séquentielle : les ménages choisissent dans un premier temps une

catégorie de véhicules (parmi six catégories), puis décident dans une seconde étape du degré d'ancienneté du véhicule (parmi quatre degrés). Plus récemment, TRAIN et WINSTON (2005) ont décodé le déclin des parts de marché aux Etats-Unis des producteurs d'automobiles locaux en exploitant les résultats d'un modèle logit mélangé (« mixed logit model »), qui décrit le choix individuel d'acquisition automobile parmi quelques 200 véhicules produits en 2000.

D'autres travaux encore viennent étoffer la revue de littérature statistique portant sur les choix qualitatifs d'automobiles : citons par exemple HOCHERMAN et al. (1983), MANNERING et MAHMASSANI (1985), BERKOVECK (1985), BERKOVECK et RUST (1985), TRAIN (1986), BROWNSTONE et al. (2000). Tous utilisent tous des modèles appartenant à la famille des modèles logit. A notre connaissance, le modèle probit multinomial MNP n'a pas encore été appliqué pour décrire les choix individuels et discrets d'acquisition automobile.

Par ailleurs, le croisement des caractéristiques des voitures considéré dans cette étude (niveau de gamme, carburant d'usage, et état à l'acquisition) pour construire des modalités de choix permet d'en limiter le nombre, et donc une manipulation statistique facilitée. Mais surtout, ces modalités nous semble plus raisonnablement correspondre à un premier plan de choix des agents que l'offre désagrégée de véhicules (ou un échantillon aléatoire) directement.

4.3 Théorie microéconomique du choix discret

4.3.1 Le choix du consommateur

Au moment d'acquérir une automobile, un individu est confronté à un problème économique : ayant pris la décision de s'équiper, il doit choisir une voiture parmi un ensemble fini de modèles concurrents que lui offre le marché. Puisque l'espace des possibilités en automobiles lui est dénombrable, le choix de l'individu est dit « discret ».

La microéconomie classique du consommateur offre un cadre théorique consistant permettant de modéliser ces comportements ces choix. La discipline suppose que l'agent est capable de donner une mesure de satisfaction retirée par la consommation de tout panier de biens, et de les comparer individuellement afin d'établir ses relations de préférence (VARIAN, 1995). La règle de choix de l'agent est ensuite soumise à sa rationalité, qui consiste en la recherche du panier de

consommation lui procurant la meilleure satisfaction possible parmi tous les paniers qui lui sont accessibles : c'est le principe de maximisation de l'utilité sous contrainte budgétaire. Le panier optimal identifié, la raison de l'agent veut qu'il concrétise son choix en le consommant.

Dans notre cas, nous considérons que l'agent peut envisager l'acquisition de tous les modèles automobiles qui lui sont proposés sans restriction de l'offre, et qu'il peut avoir accès à l'emprunt si nécessaire. Il est juste tenu de respecter son équilibre budgétaire intertemporel.

Lorsque l'on applique la théorie au contexte du choix discret d'acquisition automobile, l'individu compare tout d'abord l'ensemble des paniers de consommation optimaux accessibles, et ne possédant pour chacun que l'une des possibilités d'acquisition exclusivement. Adoptons les notations suivantes. Notons k le k -ième modèle automobile appartenant à un ensemble de choix fermé comptant p modèles : $k = 1, 2, \dots, p$. Le panier optimal d'un agent contenant k est noté $(k, B(k))$, où $B(k)$ est la quantité optimale des autres biens de consommation contrainte par la présence de k dans le panier et par le budget de l'agent. Le niveau de satisfaction (ou niveau d'utilité) que l'agent retire de la consommation de $(k, B(k))$ est donné par son application dans la fonction U à valeurs dans \mathbb{R} , dite « fonction d'utilité », et est noté $U(k, B(k))$.

En comparant les valeurs $U(k, B(k))$ pour $k = 1, 2, \dots, p$, l'agent établit ses relations de préférence entre les paniers de consommation. Pour $l \neq l'$, nous avons :

$$\begin{aligned} U(l, B(l)) > U(l', B(l')) &\Rightarrow (l, B(l)) \succ (l', B(l')), \\ U(l, B(l)) = U(l', B(l')) &\Rightarrow (l, B(l)) \sim (l', B(l')), \end{aligned} \tag{4.1}$$

où $a \succ b$ et $a \sim b$ définissent respectivement la préférence stricte et l'équivalence du panier a sur le panier b pour l'agent. Dans ces conditions, il peut ordonner les p paniers de consommation en une chaîne unique ou séquence de préférences les liant. S'il est rationnel et puisque les préférences sont transitives, alors il choisira de consommer le panier $(k, B(k))^*$ qui se trouve en bout de chaîne et qui domine au sens de ses préférences tous les autres paniers $(l', B(l'))$ pour tout $l' \neq k$. Formellement, le panier optimal de consommation $(k, B(k))^*$ est donné par :

$$(k, B(k))^* = \left((l, B(l)) : U(l, B(l)) = \max_k U(k, B(k)) \right). \tag{4.2}$$

Dans la pratique, nous n'observons pas dans le détail les étapes du raisonnement de l'agent, en particulier la façon dont il ordonne ses préférences entre les p alternatives de choix. L'agent ne dévoile que son choix final, en révélant sa préférence absolue pour le panier contenant la j -ième automobile sur tous les autres. Notons D , la variable du choix du modèle automobile de l'agent qui accompagne la consommation de $(j, B(j))$:

$$D = \left(j : (j, B(j)) = (k, B(k))^* \right). \quad (4.3)$$

Enfin, puisque la quantité $B(k)$ est fonction de k , nous pouvons redéfinir le niveau d'utilité associé à $(k, B(k))$ par Y_k^* :

$$Y_k^* = U(k, B(k)). \quad (4.4)$$

4.4 Le modèle probit multinomial (MNP)

Dans cette partie, admettons que l'on dispose d'un échantillon de N observations individuelles du comportement d'acquisition automobile, dont on considère qu'il a été généré par le schéma théorique des préférences révélées décrit dans la section précédente. Nous élargissons nos notations en pré-indiquant les variables d'intérêt par un identifiant $i = [1, 2, \dots, N]$ de l'agent présent dans l'échantillon, et notons D_i la variable polytomique du choix de l'agent i qui prend la valeur j s'il a choisi la j -ième alternative parmi un ensemble de p alternatives concurrentes et mutuellement exclusives :

$$D_i = j \text{ pour } 1 \leq j \leq p. \quad (4.5)$$

Sans perte de généralité, D_i est, dans une version multinomiale, représenté par un vecteur p -dimensionnel à composantes booléennes (0 ou 1). Soit Y_i , un vecteur $(p \times 1)$, dont la j -ième composante, Y_{ij} , prend la valeur 1 si le choix de l'agent i se porte sur l'alternative j , et dont les autres composantes, notées Y_{i-j} , prennent la valeur 0 :

$$Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ip})' \text{ avec } Y_{ij} = 1 \text{ et } Y_{i-j} = 0 \text{ si } D_i = j. \quad (4.6)$$

4.4.1 Formulation de l'utilité aléatoire

Conformément au mécanisme décisionnel décrit en section 4.3.1, l'agent i est supposé rationnel et choisira l'alternative qui lui procure le meilleur niveau d'utilité Y_{ij}^* parmi l'ensemble des p choix possibles. Le vecteur multinomial observable Y_i résulte ainsi de la discrétisation des niveaux de satisfaction $Y_i^* = (Y_{i1}^*, Y_{i2}^*, \dots, Y_{ip}^*)'$ partiellement inobservables :

$$\begin{aligned} Y_{ij} &= 1 & \text{si } Y_{ij}^* = \max_k (Y_{ik}^*), \\ Y_{ij} &= 0 & \text{si } Y_{ij}^* < \max_k (Y_{ik}^*). \end{aligned} \quad (4.7)$$

Du point de vue de l'agent i , les p niveaux d'utilité Y_{ik}^* dépendent de facteurs explicatifs exhaustivement connus, et auxquels il accorde une importance (une pondération) qui lui est individuellement propre pour effectuer son choix. Il est donc capable de donner une mesure totalement déterministe de Y_i^* .

La perspective du modélisateur est différente, tout d'abord parce qu'il ne dispose pas idéalement de l'ensemble des facteurs pouvant expliquer les choix des agents, mais aussi parce qu'il cherche généralement à évaluer la pondération moyenne des agents, ou encore le comportement moyen de choix pour des agents similaires, décrivant les mêmes facteurs explicatifs observables.

Dès lors et suivant MCFADDEN (1974), le modélisateur suppose usuellement que les niveaux d'utilité Y_{ik}^* se composent d'une partie déterministe, notée V_{ik} , et d'un terme d'erreur aléatoire v_{ik} : $Y_{ik}^* = V_{ik} + v_{ik}$. Cette relation formalise la base des modèles de choix discrets dits « à utilité aléatoire ». Elle permet d'incorporer dans le terme v_{ik} l'ensemble des variables explicatives omises par le modélisateur, ainsi que l'effet sur les niveaux d'utilité des écarts de pondération des agents avec la pondération moyenne des facteurs explicatifs considérés par le modélisateur. La partie déterministe V_{ik} peut alors être interprétée comme le niveau moyen d'utilité accordé à l'alternative k par les agents confrontés aux mêmes facteurs explicatifs observables qu'un agent i .

Généralement, V_{ik} est spécifié par une relation linéaire de facteurs explicatifs observables, comprenant les caractéristiques de l'agent i , notées \underline{X}_i , et les caractéristiques (ou attributs) de la modalité k , notées Z_k : $V_{ik} = \underline{X}_i \tilde{\beta}_k + Z_k \delta$, où $\tilde{\beta}_k$ et δ sont les pondérations, les coefficients recherchés par le modélisateur. Lorsque Z_k est invariant pour tous les agents et que \underline{X}_i contient une constante,

alors le paramètre δ n'est pas identifiable, si bien que son effet sur V_{ik} est absorbé par les constantes spécifiques aux alternatives du modèle. Dans notre cas d'étude, nous observons notamment que tous les agents sont pratiquement confrontés aux mêmes systèmes de prix et de caractéristiques physiques des modalités automobiles⁷⁶. En conséquence, seule la variabilité des caractéristiques observables \underline{X}_i entre agents nous permettra de différencier la partie V_{ik} de leurs niveaux d'utilité, et :

$$V_{ik} = \underline{X}_i \tilde{\beta}_k. \quad (4.8)$$

On aboutit à un système de régressions des variables latentes (i.e. les niveaux d'utilité) suivant : $Y_{ik}^* = \underline{X}_i \tilde{\beta}_k + v_{ik}$ pour $1 \leq k \leq p$. Pour des facilités de manipulations ultérieures, on peut réécrire le modèle matriciel sous une forme plus habituelle :

$$\begin{pmatrix} Y_{i1}^* \\ Y_{i2}^* \\ \dots \\ Y_{ip}^* \end{pmatrix} = \begin{pmatrix} \underline{X}_i & 0 & \dots & 0 \\ 0 & \underline{X}_i & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \underline{X}_i \end{pmatrix} \times \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \dots \\ \tilde{\beta}_p \end{pmatrix} + \begin{pmatrix} v_{i1} \\ v_{i2} \\ \dots \\ v_{ip} \end{pmatrix}, \quad (4.9)$$

soit :

$$Y_i^* = [I_p \otimes \underline{X}_i] \times \text{vec}(\tilde{\beta}_1, \dots, \tilde{\beta}_p) + v_i, \quad (4.10)$$

ou encore :

$$Y_i^* = X_i \tilde{\beta} + v_i. \quad (4.11)$$

Pour un individu i , sa probabilité d'avoir choisi l'alternative j , $\Pr(D_i = j)$ est égale à :

$$\Pr(D_i = j) = \Pr\left(\bigcap_{k \neq j} Y_{ik}^* \leq Y_{ij}^*\right). \quad (4.12)$$

⁷⁶ Entre 1997 et 2001 dans nos données, ces variables présentent en effet une bonne stabilité. La forte colinéarité avec les constantes spécifiques aux alternatives qui en résulte menace la bonne identification du modèle. Ce constat apparaît également à l'examen de l'évolution des indices des prix des automobiles renseignés par l'INSEE durant cette période (www.insee.fr, séries statistiques n° 063880257, 063880358, 063880459).

Cette probabilité repose sur le type supposé de distribution suivie par les aléas v_i . En prenant v_{ik} indépendants entre agents, non corrélés entre alternatives, et distribués selon une loi de Gumbel, on aboutit moyennant quelques transformations au très populaire modèle logit multinomial (MNL⁷⁷) de MCFADDEN (1974). Dans celui-ci, la probabilité pour un agent i de choisir la j -ième modalité automobile est calculée selon la formule :

$$\Pr(D_i = j) = \exp(V_{ij}) / \sum_k \exp(V_{ik}). \quad (4.13)$$

La non corrélation des aléas vaut au modèle de supposer vraie l'hypothèse IIA⁷⁸ d'indépendance des alternatives non pertinentes : on vérifie que l'« odds », ou le ratio de probabilité $O_{j-j'}^i = \Pr(D_i = j) / \Pr(D_i = j') = \exp(V_{ij} - V_{ij'})$ est indépendant des modalités k , pour $k \neq \{j, j'\}$. En pratique, la propriété n'est pas toujours souhaitable. Pour l'illustrer, l'exemple incontournable est le paradoxe bus bleu - bus rouge formulé par MCFADDEN (1974).

Supposons que l'agent i dispose initialement de deux modalités de transport pour se rendre sur son lieu de travail et qu'il ne peut en choisir qu'une : son automobile (a), ou prendre un bus bleu (b). Si le modèle logit estime qu'il est indifférent entre les deux modalités, alors : $V_{ia} = V_{ib}$. Il vient que l'odds $O_{a-b}^i = \exp(V_{ia} - V_{ib}) = 1$, et donc que $\Pr(D_i = a) = \Pr(D_i = b) = 1/2$. Imaginons maintenant qu'une société vienne élargir les possibilités de choix de l'agent en lui proposant d'aller travailler dans un bus rouge (c) ayant strictement les mêmes conditions de transport que le bus bleu. La couleur lui important peu, l'agent est indifférent entre les deux modalités de bus, et le modèle logit donne : $V_{ib} = V_{ic}$. Il en découle logiquement que $\Pr(D_i = b) = \Pr(D_i = c)$. Le résultat logiquement espéré de l'introduction de la modalité (c) est un partage de la probabilité initiale de choisir le bus bleu : $\Pr(D_i = b) = \Pr(D_i = c) = 1/4$, laissant inchangée celle de choisir l'automobile. Pourtant, en vertu de sa propriété IIA, l'introduction de la modalité (c) ne change pas la valeur du ratio $O_{a-b}^i = 1$ dans le modèle MNL, qui conclut irraisonnablement que :

⁷⁷ Sigle anglosaxon conventionnel pour « MultiNomial Logit model ».

⁷⁸ « Independence of Irrelevant Alternatives » en anglais (IIA).

$\Pr(D_i = a) = \Pr(D_i = b) = \Pr(D_i = c) = 1/3$, et que la probabilité pour l'agent de choisir le bus augmente⁷⁹ (2/3 contre 1/2 initialement)⁸⁰.

Pour résumer, si certaines modalités retenues sont de très proches substituts au regard d'autres alternatives, alors l'hypothèse IIA peut ne pas être crédible, prohibant l'usage du modèle MNL. La solution pour pallier cette défaillance consiste à lever l'hypothèse d'indépendance entre les aléas des différentes alternatives. C'est notamment ce qu'autorisent les modèles à valeur extrême généralisée, dont la famille comprend les modèles logit emboîtés. En accord avec une structure décisionnelle de choix en plusieurs étapes, ce modèle constitue un outil techniquement adapté pour la résolution du paradoxe « bus bleu - bus rouge » précédent, et plus généralement pour tout problème de choix discret résultant d'une séquence hiérarchique naturelle (GOURIEROUX, 1989, p.57). Cependant, le montage statistique de ces modèles logit⁸¹ repose sur une distribution des termes d'erreur v_i très méticuleusement choisie : la famille des lois de valeurs extrêmes est particulièrement appréciable pour dériver une formulation simple des probabilités de choix discrets. Or, que penser d'une telle famille de lois statistiques, à la fois non normales, et dont les espérances ne sont pas nulles pour décrire des termes d'erreurs? Même si techniquement cela n'est pas fondamental au regard des quantités d'intérêt que sont les différences d'utilité⁸², le caractère purement *ad hoc* de choisir une loi de valeurs extrêmes pour caractériser le vecteur v_i peut laisser le praticien habitué aux méthodes d'estimation linéaires insatisfait pour paramétrer le système de régressions (4.9). Historiquement, les modèles logit ont été introduits comme substituts ou approximations des modèles probit, plus complexes (GOURIEROUX, 1989, p.29). Mais surtout et comme conséquence du théorème central limite, la somme de tous les effets inobservés par le modélisateur tendrait davantage à être distribuée selon une loi normale.

⁷⁹ Si le modèle MNL était vraiment approprié pour décrire le choix du mode de transport *dans ces conditions*, alors une politique publique efficace visant à réduire l'usage de l'automobile serait de repeindre la flotte des bus en différentes couleurs...

⁸⁰ Un exemple plus ancien illustrant ce problème est donné par DEBREU (1960), commentant les travaux de LUCE (1959). Dans celui-ci, un amateur de musique classique a tout d'abord le choix entre écouter un enregistrement de la 8^{ème} symphonie de Beethoven dirigée par un chef d'orchestre ou un quartet de Debussy, puis l'espace des choix est augmenté par un nouvel enregistrement de la 8^{ème} de Beethoven, mais dirigée par un autre chef d'orchestre.

⁸¹ La formulation de ces modèles n'est pas abordée. Nous renvoyons le lecteur à l'ouvrage de TRAIN (2002) pour leur description.

⁸² Nous le verrons dans le paragraphe suivant, à propos de l'identification du modèle.

Toutes ces motivations nous conduisent à préférer l'hypothèse que $v_i = (v_{i1}, \dots, v_{ip})'$ est normalement et indépendamment distribué : $v_i \sim \mathcal{N}_p(0, \Sigma^*)$, avec Σ^* une matrice générale de covariances $p \times p$ sans restriction d'indépendance des aléas entre les alternatives de choix. Si les niveaux d'utilité Y_i^* étaient observés, cette hypothèse formaliserait un modèle de régressions empilées⁸³. Appliqué au contexte des choix discrets, ce paramétrage des aléas caractérise le modèle probit multinomial, noté MNP (HAUSMAN et WISE, 1978 ; BOLDUC, 1999).

4.4.2 Identification du modèle MNP

Le modèle spécifié n'est pas formellement identifiable. Notamment, l'ajout d'un scalaire quelconque dans chaque niveau d'utilité apporté par les alternatives laisse inchangée la structure des choix de l'agent. Ce problème de calage des paramètres (« location invariance ») est résolu en différenciant le système des régressions des variables latentes avec l'une d'entre elles, désignant une alternative de référence (ici l'alternative p) :

$$W_{ik} = Y_{ik}^* - Y_{ip}^* = \underline{X}_i (\tilde{\beta}_k - \tilde{\beta}_p) + (v_{ik} - v_{ip}). \quad (4.14)$$

Ce passage aux différences permet d'observer que l'on ne pourra, à ce niveau, identifier que les vecteurs de paramètres $\beta_k = (\tilde{\beta}_k - \tilde{\beta}_p)$. Intuitivement, on comprend aisément que les p niveaux d'utilité n'apportent pas plus d'information que les $p-1$ différences d'utilité, relatives à celle du choix p . Il est donc équivalent d'écrire le modèle MNP sous cette forme, pour $1 \leq k \leq p-1$:

$$\begin{aligned} W_{ik} &= \underline{X}_i \beta_k + \varepsilon_{ik}, \\ W_{ip} &= 0, \\ D_i &= j \text{ si } W_{ij} = \max_k (W_{ik}) \text{ et } W_{ij} > 0, \\ D_i &= p \text{ si } \max_k (W_{ik}) < 0. \end{aligned} \quad (4.15)$$

Matriciellement, pour $1 \leq k \leq p-1$:

⁸³ Plus connu dans sa dénomination anglosaxonne : « Seemingly Unrelated Regressions (SUR) model ».

$$\begin{pmatrix} W_{i1} \\ W_{i2} \\ \dots \\ W_{ip-1} \end{pmatrix} = \begin{pmatrix} \underline{X}_i & 0 & \dots & 0 \\ 0 & \underline{X}_i & \dots & 0 \\ 0 & \dots & \underline{X}_i & 0 \\ 0 & \dots & 0 & \underline{X}_i \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{ip-1} \end{pmatrix}, \quad (4.16)$$

$$\text{soit :} \quad W_i = (I_{p-1} \otimes \underline{X}_i) \times \text{vec}(\beta_1, \dots, \beta_{p-1}) + \varepsilon_i, \quad (4.17)$$

$$\text{ou encore :} \quad W_i = X_i \beta + \varepsilon_i, \quad (4.18)$$

$$\begin{aligned} \varepsilon_i &= (\varepsilon_{i1}, \dots, \varepsilon_{ip-1})', \\ \text{avec :} \quad \varepsilon_i &\sim \mathcal{N}(0, \Omega), \\ \Omega &= [I_{p-1} \sim -e_{p-1}] \Sigma^* [I_{p-1} \sim -e_{p-1}]', \end{aligned} \quad (4.19)$$

où le signe « \sim » désigne l'opérateur de concaténation horizontale, et e_{p-1} désigne une matrice colonne composée de 1 sur $p - 1$ lignes. Si le problème du calage des paramètres est résolu par la différenciation, l'identification des paramètres du modèle n'est pas encore acquise. Notamment, la multiplication par un scalaire positif non nul ($c > 0$) de l'ensemble des utilités relatives W_{ik} laisse également inchangée la structure des préférences des agents entre les alternatives :

$$D(W_i) = D(cW_i) \rightarrow D(X_i \beta + \varepsilon_i) = D(X_i c \beta + c \varepsilon_i). \quad (4.20)$$

Ce problème d'échelle peut être surmonté en fixant l'une des composantes de β , ce qui demande la connaissance *a priori* de son signe. Plus habituellement, comme dans le modèle probit binaire classique, on fixe conventionnellement à 1 une covariance de Ω . En choisissant arbitrairement celle de la première modalité, cela revient à fixer $c = 1/\sqrt{\Omega_{11}}$. Notons ainsi W_i^* les différentiels d'utilité « normalisés », puis β^* et Ω^* les paramètres identifiables, définis par :

$$W_i^* = W_i / \sqrt{\Omega_{11}}, \quad \beta^* = \beta / \sqrt{\Omega_{11}}, \quad \Omega^* = \Omega / \Omega_{11}, \quad (4.21)$$

avec $\Omega_{11}^* = 1$. L'identification usuelle du modèle achevée, on peut facilement écrire la vraisemblance du modèle MNP en fonction de β^* et Ω^* :

$$L(Y | \beta^*, \Omega^*, X) = \prod_{i=1}^N L(Y_i | \beta^*, \Omega^*, X_i), \quad (4.22)$$

$$L(Y_i | \beta^*, \Omega^*, X_i) = \int_{Z_i} (2\pi)^{-0.5(p-1)} |\Omega^*|^{-0.5} \exp\left(-0.5(W_i^* - X_i\beta^*)' (\Omega^*)^{-1} (W_i^* - X_i\beta^*)\right) dZ_i \quad (4.23)$$

$$\text{avec : } Z_i = \begin{cases} Z_i : \left(\bigcap_{k \neq j} (W_{ij}^* > W_{ik}^*) \right) \cap (W_{ij}^* > 0) \text{ si } D_i = j \text{ et } D_i \neq p \\ Z_i : \bigcap_k (W_{ik}^* < 0) \text{ si } D_i = p \end{cases}. \quad (4.24)$$

L'estimateur de β^* et Ω^* par maximum de vraisemblance implique donc de maximiser la fonction $L(Y | \beta^*, \Omega^*, X)$. Cette procédure demande la résolution individuelle de N intégrales multiples de dimension $p-1$, particulièrement difficiles (voire impossibles) à évaluer lorsque p est supérieur à trois. Pour ce motif, l'estimation du modèle MNP s'est très souvent effectuée à l'aide des méthodes de simulations. LERMAN et MANSKI (1981) ont par exemple proposé une approche par maximum de vraisemblance simulée (MVS), où les probabilités de choix entre alternatives sont obtenues via un simulateur de fréquences brut⁸⁴ (« Crude Frequency Simulator », CFS). MCFADDEN (1989) a étendu la méthode des moments simulés pour estimer des probabilités de choix, en utilisant une version lissée du simulateur de fréquences brut (« Kernel Smoothed Frequency Simulator »). Ces deux méthodes précédentes ont ensuite exploité avec bénéfice le simulateur GHK (pour Geweke-Hajivassiliou-Keane ; GEWEKE, 1991 ; BORSH-SUPAN et HAJIVASSILIOU, 1993) pour calculer les probabilités. Dans un comparatif des différentes méthodes d'ajustement du modèle MNP, GEWEKE et al. (1994) ont notamment conclu aux bonnes propriétés de l'approche bayésienne, adoptée dans ce document.

4.4.3 Analyse Bayésienne du MNP

Nous reprenons ici l'approche bayésienne du MNP de McCULLOCH et ROSSI (1994). Dans celle-ci, l'analyse du modèle passe par la spécification de distributions

⁸⁴ Que nous utiliserons plus tard pour l'exploitation des résultats d'estimation.

a priori pour les paramètres β et $G = \Omega^{-1}$. Pratiquement, on choisit des distributions indépendantes et qui sont, nous le verrons, conjuguées avec la vraisemblance de W :

$$\beta \sim \mathcal{N}(b, A), \quad P(\beta) \propto |A|^{-0.5} \exp\left(-0.5(\beta - b)' A^{-1} (\beta - b)\right), \quad (4.25)$$

$$\Omega^{-1} \sim \mathcal{W}_{p-1}(v, V), \quad P(\Omega^{-1}) = P(G) \propto |G|^{(v-p)/2} \exp\left(-0.5 \text{tr}(GV^{-1})\right), \quad (4.26)$$

où \mathcal{W} est le symbole conventionnel de la distribution de Wishart⁸⁵. Le produit de la fonction de vraisemblance du modèle MNP avec ces distributions *a priori* est proportionnel à la distribution *a posteriori* des paramètres (CHIB, 1996) :

$$P(\beta, G \mid X, Y) \propto L(Y \mid \beta, G, X) \times P(\beta, G). \quad (4.27)$$

Relativement à l'information apportée par la vraisemblance sur les paramètres, le caractère plus ou moins précis des distributions *a priori* détermine leur degré d'influence sur la distribution *a posteriori*. Comme nous l'avons déjà évoqué dans le chapitre précédent, le passage de la distribution *a priori* vers la distribution *a posteriori* peut directement être interprété comme un processus de révision bayésien des croyances : le modélisateur reflète ses opinions et leur force en premier lieu dans les distributions *a priori* des paramètres : $P(\beta)$ et $P(G)$, puis sous l'éclairage de l'information contenue par les données (X, Y) , il révisé ses croyances initiales en formulant une distribution *a posteriori* : $P(\beta, G \mid X, Y)$. Dans notre étude, il convient d'imposer des distributions *a priori* très diffuses pour que la distribution *a posteriori* épouse l'allure de la fonction de vraisemblance $L(Y \mid \beta, G, X)$. Pour $P(\beta)$ notamment, il suffit de choisir A avec des éléments diagonaux très grands, et pour $P(G)$, il convient de choisir une valeur pour v très petite par rapport à la taille N de l'échantillon et telle que $v > p$ (McCULLOCH et ROSSI, 1994).

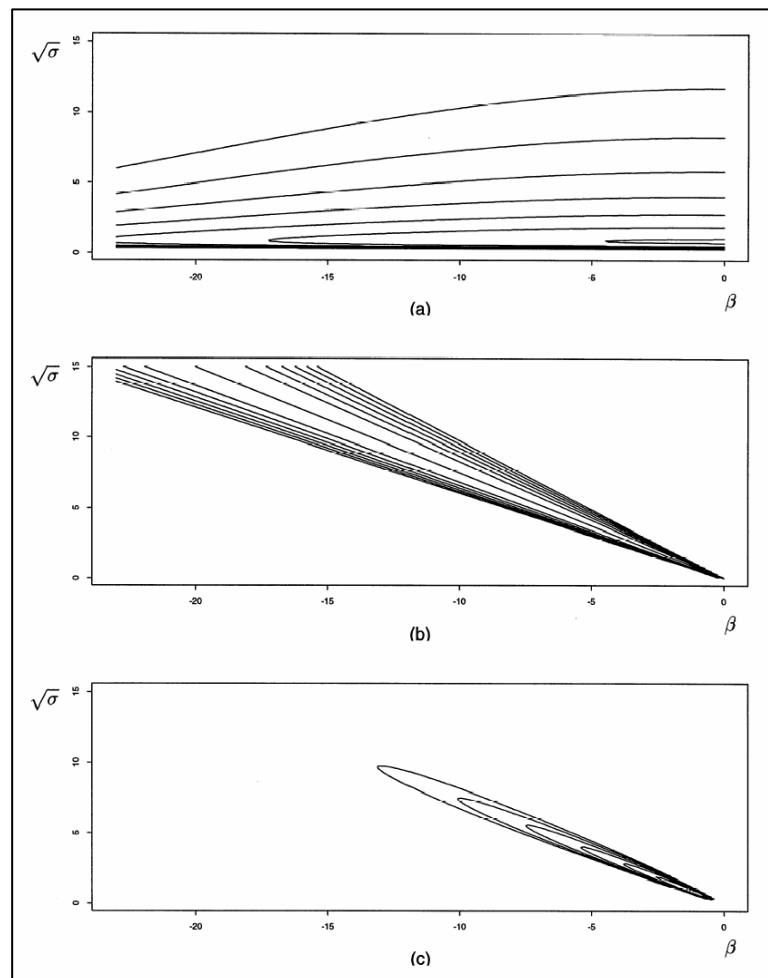
Remarquons que la spécification *a priori* de G n'impose pas la restriction d'identification $(G^{-1})_{1,1} = \Omega_{1,1} = 1$. Bien que la partie informative (la fonction de

⁸⁵ Nommée ainsi en l'honneur du statisticien John Wishart (1898-1956). Soit une variable aléatoire U qui suit une $\mathcal{N}_T(0, \Gamma)$, et U_i la i -ième réalisation de U (une matrice colonne avec T lignes), alors la matrice $\sum_{i=1}^n U_i U_i'$ est distribuée selon une loi de Wishart d'hyperparamètre Γ avec n degrés de liberté, notée $\mathcal{W}_T(n, \Gamma)$.

vraisemblance) ne permette pas à elle seule l'identification, celle-ci est dans l'approche bayésienne obtenue par l'usage de distributions *a priori* « propres » (continues et définies sur l'espace des valeurs possibles des paramètres et dont l'intégration sur cette espace sous la fonction de densité est égale à 1). Il en résulte une distribution *a posteriori* également propre.

Dans la figure 4.1, NOBILE (1998) illustre pourquoi la démarche bayésienne n'a pas besoin de poser l'*a priori* $\Omega_{1,1} = 1$ en prenant l'exemple d'un modèle probit binaire avec un paramètre d'intérêt β et un aléa univarié normal de variance σ .

FIGURE 4.1 : Exemple du passage de la distribution a priori à la distribution a posteriori des paramètres dans un modèle probit binaire



Légende : (a) distribution *a priori*, (b) allure de la vraisemblance, (c) distribution *a posteriori*. Source : NOBILE (1998).

Alors que la fonction de vraisemblance est constante pour tout ratio $\beta/\sqrt{\sigma}$ (figure 4.1, panel b), elle ne permet pas d'identifier β sans poser de restriction sur σ (usuellement fixé à 1). Mais en spécifiant une distribution *a priori* propre pour β et $\sqrt{\sigma}$ (figure 4.1, panel a), il résulte de sa combinaison avec la fonction de vraisemblance une distribution *a posteriori* (figure 4.1, panel c) propre faisant apparaître, à la place des lignes continues, des ellipses d'autant plus allongées dans la direction de la vraisemblance que les distributions *a priori* sur β et σ^{-1} sont faiblement informatives.

Néanmoins, la masse distributionnelle *a posteriori* est concentrée sur l'espace des paramètres situé sous les ellipses de densité les plus élevées, dont la longueur n'est déterminée que par la variance des distributions *a priori* puisque la fonction de vraisemblance est linéairement constante. Ainsi, et en se ramenant à notre cas général du probit multinomial, la masse des échantillonnages dans la distribution *a posteriori* doit se concentrer sur un ensemble de points $(c\beta, c^2\Omega, c > 0)$ plutôt que sur un point unique. Il est donc nécessaire pour éliminer l'effet d'échelle c de ne pratiquer l'inférence que sur la forme identifiée des paramètres (échantillonnés) $\beta^* = \beta/\Omega_{11}^{0.5}$ et $\Omega^* = \Omega/\Omega_{11}^{0.5}$, dont dépend uniquement la fonction de vraisemblance (4.22).

4.5 La méthode MCMC

Comme nous l'avons évoqué, l'évaluation de la seule fonction de vraisemblance (4.22) est très difficile. Les méthodes numériques pour l'analyse bayésienne s'avèrent également très décourageantes. L'approche par simulations de Monte Carlo, consistant à faire des tirages dans une distribution, permet de s'affranchir du calcul complexe d'intégrales par une simple exploitation statistique de ces tirages. Nous utilisons pour l'estimation une méthode de Monte Carlo par chaînes de Markov (MCMC) : l'échantillonnage de Gibbs (GEMAN et GEMAN, 1984) avec augmentation de données (TANNER et WONG, 1987) est mis à l'œuvre pour caractériser $P(\beta^*, G^* | X, Y)$. L'intérêt de l'algorithme de l'échantillonnage de Gibbs repose sur le remarquable résultat que des tirages itératifs et récursifs dans des lois conditionnelles des paramètres donnent asymptotiquement lieu à une séquence de réalisations qui converge en distribution vers la loi jointe des paramètres. L'étape d'augmentation de données est une méthode de complétion qui

consiste à étendre l'espace des paramètres β et G avec le vecteur des variables latentes W , et qui nous affranchit de toute intégration.

4.5.1 Echantillonnage de Gibbs

Rappelons brièvement le principe de fonctionnement de l'échantillonneur de Gibbs. Soit θ , le vecteur de paramètres partitionnable en θ_1 et θ_2 .

Etape 1 : L'algorithme d'échantillonnage est initialisé en $\theta_1^{(0)}$ et $\theta_2^{(0)}$.

Etape 2 : On échantillonne successivement dans les distributions conditionnelles suivantes : $\theta_1^{(g+1)}$ tiré dans $f(\theta_1 | \theta_2 = \theta_2^{(g)})$, $\theta_2^{(g+1)}$ tiré dans $f(\theta_2 | \theta_1 = \theta_1^{(g+1)})$.

Etape 3 : $g \leftarrow g + 1$, retour à l'étape 2.

Il en résulte $\overrightarrow{\theta^{(h)}}$, une séquence de $\theta^{(h)}$ qui converge en distribution vers la loi jointe $f(\theta_1, \theta_2)$.

Adaptons maintenant cet algorithme d'échantillonnage à nos notations. Soit θ_+ , l'ensemble des paramètres augmenté des W_{ij} , décomposable en $\theta_+ = \{(W_{11}, \dots, W_{1p-1}), \dots, (W_{N1}, \dots, W_{Np-1}), \beta, G\}$.

Etape 1 : L'algorithme d'échantillonnage est initialisé en $\theta_+^{(0)} = \{W^{(0)}, \beta^{(0)}, G^{(0)}\}$.

Etape 2 : On échantillonne successivement dans les distributions conditionnelles suivantes :

$W_{ik}^{(g+1)}$ tiré dans $P(W_{ik} | W_{i < k} = W_{i < k}^{(g+1)}, W_{i > k} = W_{i > k}^{(g)}, \beta = \beta^{(g)}, G = G^{(g)}, X_i, Y_i)$,

pour chaque k de 1 jusque $p-1$ et tout i de 1 jusque N , afin de créer

$W^{(g+1)}$, où $W_{i < k}^{(g+1)} = (W_{i1}^{(g+1)}, \dots, W_{ik-1}^{(g+1)})$ et où $W_{i > k}^{(g)} = (W_{ik+1}^{(g)}, \dots, W_{ip-1}^{(g)})$,

$\beta^{(g+1)}$ tiré dans $P(\beta | W = W^{(g+1)}, G = G^{(g)}, X, Y)$,

$G^{(g+1)}$ tiré dans $P(G | W = W^{(g+1)}, \beta = \beta^{(g+1)}, X, Y)$.

Etape 3 : $g \leftarrow g + 1$, retour à l'étape 2.

Il en résulte $\overrightarrow{\theta_+^{(h)}}$, une suite (une chaîne de Markov) de $\theta_+^{(h)}$ qui converge en distribution vers la loi jointe $P(W, \beta, G | X, Y)$ lorsque h tend vers l'infini. Ecarter $W^{(h)}$ des échantillonnages $\theta_+^{(h)}$ a pour effet de marginaliser la distribution des paramètres β et G : il en résulte $\overrightarrow{\theta^{(h)}}$ une suite de $\theta^{(h)}$ qui converge en distribution vers $P(\beta, G | X, Y)$, une première distribution d'intérêt.

Spécifiquement pour i , l'échantillonnage de son vecteur d'utilités relatives W_i directement dans une distribution multivariée et tronquée aurait pu être conduit avec une méthode d'« acceptation-rejet » décrite par ALBERT et CHIB (1993). Le désavantage de cette méthode est qu'il faut parfois effectuer plusieurs tirages avant d'accepter celui qui tombera dans la bonne région : celle impliquée par le choix D_i . C'est pourquoi nous avons préféré retenir la démarche de MCCULLOCH et ROSSI (1994), en partitionnant le vecteur des variables latentes W_i en variables univariées W_{ik} dans l'échantillonneur de Gibbs, et pour lesquelles l'échantillonnage conforme à D_i est plus commode. Pour mettre en application l'échantillonnage de Gibbs augmenté, il est maintenant nécessaire de dériver et d'explicitier les distributions conditionnelles *a posteriori* de chaque élément de θ_+ .

4.5.2 Les distributions conditionnelles a posteriori

Formellement, les distributions conditionnelles dans l'étape 2 de l'échantillonneur de Gibbs augmenté sont obtenues à partir de la relation de proportionnalité (4.27) : $P(\beta, G | X, Y) \propto L(Y | \beta, G, X) \times P(\beta) \times P(G)$. Avec l'augmentation de données, nous obtenons cette relation :

$$P(W, \beta, G | Y, X) \propto L(Y, W | \beta, G, X) \times P(\beta) \times P(G). \quad (4.28)$$

La décomposition conditionnelle de $L(Y, W | \beta, G, X)$ donne :

$$P(W, \beta, G | Y, X) \propto L(Y | W, \beta, G, X) \times L(W | \beta, G, X) \times P(\beta) \times P(G). \quad (4.29)$$

Puisque Y est le résultat de la discrétisation de W (système (4.15)), il vient que $L(Y | W, \beta, G, X) = L(Y | W)$. $L(Y | W)$ est une fonction indicatrice qui vaut 1 si W engendre effectivement Y , 0 sinon. Ainsi, nous avons la relation :

$$P(W, \beta, G | Y, X) \propto L(Y | W) \times L(W | \beta, G, X) \times P(\beta) \times P(G). \quad (4.30)$$

Les distributions conditionnelles qui en découlent sont :

$$P(W_i | \beta, G, Y_i, X_i) \propto L(Y_i | W_i) \times L(W_i | \beta, G, X_i), \quad (4.31)$$

$$\text{avec : } L(W_i | \beta, G, X_i) = (2\pi)^{-0.5(p-1)} |\Omega|^{-0.5} \exp\left(-0.5(W_i - X_i\beta)' \Omega^{-1} (W_i - X_i\beta)\right), \quad (4.32)$$

où $L(Y_i | W_i)$ est la fonction indicatrice précédente spécifique à l'observation i , et où $L(W_i | \beta, G, X_i)$ est la fonction de vraisemblance de W_i . Puisque $W_i | \beta, G, X_i$ suit une loi normale (cf. (4.18)), il vient que $W_i | \beta, G, Y_i, X_i$ suit une distribution normale tronquée de dimension $p - 1$. En considérant particulièrement l'alternative k , nous obtenons :

$$P(W_{ik} | W_{i-k}, \beta, G, Y_i, X_i) \propto L(Y_i | W_{ik}, W_{i-k}) \times L(W_{ik} | W_{i-k}, \beta, G, X_i), \quad (4.33)$$

où $W_{ik} | W_{i-k}, \beta, G, Y_i, X_i$ est cette fois distribué selon une loi normale tronquée unidimensionnelle, dans laquelle il est beaucoup plus simple d'échantillonner que dans $W_i | \beta, G, Y_i, X_i$ précédemment (voir GEWEKE, 1991) ;

$$P(\beta | W, G, Y, X) \propto L(W | \beta, G, X) \times P(\beta), \quad (4.34)$$

$$P(G | W, \beta, Y, X) \propto L(W | \beta, G, X) \times P(G). \quad (4.35)$$

Nous allons vérifier la conjugance⁸⁶ de $L(W | \beta, G, X)$ avec $P(\beta)$ et $P(G)$: il viendra que $\beta | W, G, Y, X$ suit une loi normale, et que $G | W, \beta, Y, X$ suit une loi de Wishart, comme leurs distributions *a priori* respectives.

Explicitons maintenant les distributions (4.33), (4.34) et (4.35). Dans l'étape 2 de l'échantillonneur de Gibbs augmenté, nous sommes tout d'abord intéressés aux tirages des variables latentes $W_{ik} | W_{i-k}, \beta, G, Y_i, X_i$ qui suivent une loi univariée normale tronquée par la région conforme à la décision de l'agent i , Y_i : $W_{ik} | W_{i-k}, \beta, G, Y_i, X_i \sim \mathcal{NT}_{Y_i}(m_{ik}, r_{ik})$. Les moyennes et variances conditionnelles, m_{ik} et r_{ik} , peuvent facilement être déduites après un réagencement de la matrice $\Omega = G^{-1}$ en trois sous-ensembles. Notons Ω_{-k-k} , la matrice Ω dont on a retiré la k -ième ligne et la k -ième colonne ; Ω_{-kk} , le vecteur de la k -ième colonne de Ω

⁸⁶ Cf. note 67 pour un rappel de la propriété de conjugance.

amputé de Ω_{kk} ; Ω_{kk} , la variance de ε_{ik} , située à l'intersection de la k -ième ligne et de la k -ième colonne dans Ω . Schématiquement et après avoir déplacé par convenance l'alternative k en dernière position :

$$\Omega = \begin{pmatrix} \begin{pmatrix} & & \\ & \Omega_{-k-k} & \\ & & \end{pmatrix} & \begin{pmatrix} \\ \Omega_{-kk} \\ \end{pmatrix} \\ \begin{pmatrix} \\ \Omega'_{-kk} & \end{pmatrix} & (\Omega_{kk}) \end{pmatrix}. \quad (4.36)$$

On dérive la variance conditionnelle de $\varepsilon_{ik} | \varepsilon_{i-k}$ à l'aide des formules usuelles : $Var(\varepsilon_{ik} | \varepsilon_{i-k}) = r_{kk} = \Omega_{kk} - \Omega'_{-kk} (\Omega_{-k-k})^{-1} \Omega_{-kk}$. De même pour l'espérance de $\varepsilon_{ik} | \varepsilon_{i-k}$: $E(\varepsilon_{ik} | \varepsilon_{i-k}) = S_{ik} = \Omega'_{-kk} (\Omega_{-k-k})^{-1} (W_{i-k} - X_{i-k} \beta)$, où W_{i-k} est le vecteur colonne W_i amputé de W_{ik} ; et où X_{i-k} est la matrice des caractéristiques X_i amputée de la k -ième ligne. Enfin, on calcule l'espérance de $W_{ik} | W_{i-k}$: $E(W_{ik} | W_{i-k}) = m_{ik} = X_{ik} \beta + S_{ik}$, où X_{ik} est la k -ième ligne de X_i . Il convient ensuite d'échantillonner W_{ik} dans une distribution $\mathcal{NT}_{Y_i}(m_{ik}, r_{kk})$, où les troncatures sont imposées par le choix de l'agent i :

$$W_{ik} | W_{i-k}, \beta, G, Y_i, X_i \sim \mathcal{NT}_{Y_i}(m_{ik}, r_{kk}). \quad (4.37)$$

Si $D_i = j$, et $j \neq p$, alors :

W_{ij} est tiré dans l'intervalle $]\max(0, W_{i-j}); +\infty[$,

W_{ik} est tiré dans l'intervalle $]-\infty; W_{ij}[$ pour $k \neq j$ et $1 \leq k \leq p-1$.

Si $D_i = p$, alors :

W_{ik} est tiré dans l'intervalle $]-\infty; 0[$ pour $1 \leq k \leq p-1$.

Pour effectuer directement des tirages dans la bonne zone d'échantillonnage, nous avons utilisé la méthode par inversion de la fonction de répartition (DEVROYE, 1986).

Après avoir simulé W , l'étape 2 de l'échantillonneur de Gibbs augmenté concerne l'échantillonnage du vecteur de paramètres β dans sa distribution *a posteriori*, conditionnellement à W , G , Y et X : $\beta | W, G, Y, X$. Il est particulièrement accommodant de transformer les vecteurs X et W avec la décomposition de Cholesky de G , pour se ramener au cas d'une régression avec des

erreurs indépendantes et identiquement distribuées entre les alternatives. Soit C cette transformation, telle que $G = CC'$. Le modèle transformé devient :

$$\begin{aligned} C'W_i &= C'X_i\beta + C'\varepsilon_i, \\ W_i^C &= X_i^C\beta + \varepsilon_i^C, \\ \varepsilon_i^C &\sim N(0, I_{p-1}). \end{aligned} \quad (4.38)$$

On déduit la distribution conditionnelle *a posteriori* de β , $P(\beta | W, G, Y, X)$, en effectuant le produit en (4.34) de $L(W | \beta, G, X) = L(W^C | \beta, X^C)$ avec $P(\beta)$, écrits comme des fonctions de β :

$$L(W^C | \beta, X^C) \propto \exp \left[-0.5 (\beta - (X^{C'}X^C)^{-1} (X^{C'}W^C))' \dots \dots (X^{C'}X^C) (\beta - (X^{C'}X^C)^{-1} (X^{C'}W^C)) \right], \quad (4.39)$$

$$P(\beta) \propto \exp(-0.5 (\beta - b)' A^{-1} (\beta - b)). \quad (4.40)$$

En utilisant la formule matricielle générale :

$$(\beta - \pi_1)' \Psi_1 (\beta - \pi_1) + (\beta - \pi_2)' \Psi_2 (\beta - \pi_2) = (\beta - \pi_3)' (\Psi_1 + \Psi_2) (\beta - \pi_3), \text{ où } \pi_3 = (\Psi_1 + \Psi_2)^{-1} (\Psi_1 \pi_1 + \Psi_2 \pi_2), \text{ il résulte du produit de (4.39) avec (4.40) que :}$$

$$P(\beta | W, G, Y, X) \propto \exp(-0.5 (\beta - b_1)' B^{-1} (\beta - b_1)), \quad (4.41)$$

avec : $B = (A^{-1} + X^{C'}X^C)^{-1}$ et $b_1 = B(A^{-1}b + X^{C'}W^C)$. Ainsi $\beta | W, G, Y, X$ suit une loi normale de paramètres b_1 et B :

$$\beta | W, G, Y, X \sim \mathcal{N}(b_1, B). \quad (4.42)$$

Enfin, après avoir simulé W et β , l'étape 2 de l'échantillonneur de Gibbs augmenté se termine par l'échantillonnage de G dans sa distribution conditionnelle *a posteriori*, $P(G | W, \beta, Y, X)$. Celui-ci s'obtient par le produit en (4.35) de $L(W | \beta, G, X)$ avec $P(G)$, écrits comme des fonctions de G :

$$P(G) \propto |G|^{(v-p)/2} \exp(-0.5 \text{tr}(GV^{-1})), \quad (4.43)$$

$$L(W | \beta, G, X) \propto |G|^{N/2} \exp\left(-0.5 \operatorname{tr}\left(\sum_{i=1}^N (W_i - X_i \beta)(W_i - X_i \beta)'\right) G\right). \quad (4.44)$$

Comme résultat du produit de (4.43) avec (4.44), il vient que :

$$P(G | W, \beta, Y, X) \propto |G|^{(v+N-p)/2} \exp\left(-0.5 \operatorname{tr}(V_1^{-1} G)\right), \quad (4.45)$$

avec $V_1 = \left(V^{-1} + \sum_{i=1}^N (W_i - X_i \beta)(W_i - X_i \beta)'\right)^{-1}$ et N la taille de l'échantillon. Ainsi, $G | W, \beta, Y, X$ suit une distribution de Wishart de paramètres $v + N$ et V_1 :

$$G | W, \beta, Y, X \sim \mathcal{W}_{p-1}(v + N, V_1). \quad (4.46)$$

La simulation par tirages d'une distribution de Wishart peut être aisément conduite grâce à la décomposition de Bartlett⁸⁷. L'étape d'augmentation de données est venue considérablement simplifier l'analyse bayésienne du modèle MNP. Celle-ci, en ajoutant l'ensemble des variables latentes W à l'espace des paramètres dans l'échantillonneur de Gibbs, transforme le modèle MNP en modèle bayésien linéaire qui ne nécessite plus aucune intégration complexe.

Jusqu'à présent, ni $P(G)$, ni les tirages dans $G | W, \beta, Y, X$ dans une distribution de Wishart n'ont eu à imposer la restriction d'identification usuelle, de fixer un élément Ω_{jj} , telle que $\Omega_{11} = (G^{-1})_{11} = 1$, comme le réclame la seule fonction de vraisemblance (section 4.4.2). Cela n'empêche pas au modèle MNP bayésien d'être identifiable dans la mesure où l'usage de distributions *a priori* propres pour β et G donne lieu à une distribution *a posteriori* de $\beta, G | Y, X$ également propre (section 4.4.3). Ceci évite la dérive des échantillonnages, et préserve ainsi la récurrence de la chaîne $\bar{\theta}^h$ (McCulloch et Rossi, 1994). Néanmoins, les écarts entre $P(\beta, G | Y, X)$ et $P(c\beta, c^{-2}G | Y, X)$ sont uniquement le fait des différences entre $P(\beta, G)$ et $P(c\beta, c^{-2}G)$, puisque la vraisemblance est constante pour tous les ensembles de la forme $\{c\beta, c^{-2}G; \forall c > 0\}$: $L(Y | \beta, G, X) = L(Y | c\beta, c^{-2}G, X)$, $\forall c > 0$. En conséquence, la distribution de la chaîne $\bar{\theta}^h$ doit asymptotiquement se concentrer sur un ensemble de $\{c\beta, c^{-2}G; c > 0\}$ plutôt que sur un point unique $\{\beta, G\}$. Au final, il convient donc

⁸⁷ Par exemple : prenons S une matrice $(T \times T)$ triangulaire inférieure telle que l'on échantillonne ses éléments (non nuls) dans les distributions suivantes : $s_{ii} \sim \sqrt{\chi_{r-i+1}^2}$, $s_{ij} \sim \mathcal{N}(0, 1)$. Alors la matrice $P = SS'$ est un tirage aléatoire dans une $\mathcal{W}_r(r, I)$. Soit K , la décomposition de Cholesky de R , de dimension $(T \times T)$ telle que $R = KK'$, alors la matrice KPK' est un tirage aléatoire dans une $\mathcal{W}_r(r, R)$.

de se ramener à une forme identifiée des paramètres pour la vraisemblance, et de ne pratiquer l'inférence que sur des quantités normalisées (4.21), telles que : $\beta^* = \beta / \sqrt{\Omega_{11}}$, $\Omega^* = \Omega / \Omega_{11} = G^{-1} / (G^{-1})_{11}$. Ainsi, pour chaque $\theta^{(h)} = \{\beta^{(h)}, G^{(h)}\}$, on calcule $\theta^{*(h)} = \{\beta^{*(h)}, \Omega^{*(h)}\}$, dont la succession lorsque $h \rightarrow \infty$ forme une chaîne $\overrightarrow{\theta^{*(h)}}$ qui tend en distribution vers $P(\beta^*, \Omega^* | Y, X)$, notre ultime distribution d'intérêt.

4.5.3 Version hybride de l'échantillonnage de Gibbs

NOBILE (1998) a proposé d'intégrer une étape supplémentaire aux cycles d'échantillonnage de Gibbs qui peut améliorer sa capacité à explorer la distribution jointe *a posteriori* $P(\beta, G | Y, X)$. Plus précisément, l'auteur intègre à l'algorithme d'échantillonnage de Gibbs la possibilité de changer l'échelle de $\theta^{(h)} = \{\beta^{(h)}, G^{(h)}\}$ en $\theta_M^{(h)} = \{c\beta^{(h)}, c^{-2}G^{(h)}\}$ en tirant c dans une distribution choisie, et validée, ou non, par une étape « Metropolis ». En effet, il n'est d'aucune importance entre deux échantillonnages de Gibbs de changer l'échelle du point de vue des paramètres normalisés $\theta^{*(h)} = \{\beta^{*(h)}, \Omega^{*(h)}\}$ qui restent constants quel que soit $c > 0$. Mais cela peut favoriser un parcours d'échantillonnage plus rapide de l'espace des paramètres sous la distribution *a posteriori* $P(\beta, G | X, Y)$ et une convergence accrue. Par ailleurs, les travaux de NOBILE (1998) sur la qualité d'ajustement du modèle MNP bayésien à partir de données générées suggèrent que l'échantillonneur de Gibbs seul tend à sous-évaluer la valeur absolue des covariances de Ω^* , alors que l'ajout de l'étape Metropolis semble améliorer quelque peu sa capacité à les révéler. Détaillons tout d'abord les étapes de cet algorithme Metropolis-Hastings (HASTINGS, 1970) :

Soit $f(\cdot)$ une distribution cible.

Soit $q(\cdot)$ une distribution instrumentale qui couvre l'espace de distribution de $f(\cdot)$.

On génère $y^{(t)}$ à partir de $q(y | x^{(t)})$,

on accepte $x^{(t+1)} = y^{(t)}$ avec une probabilité $\rho(x^{(t)}, y^{(t)})$,

on choisi $x^{(t+1)} = x^{(t)}$ avec la probabilité $1 - \rho(x^{(t)}, y^{(t)})$,

$$\text{avec } \rho(x^{(t)}, y^{(t)}) = \min \left(\frac{f(y^{(t)})}{f(x^{(t)})} \times \frac{q(x^{(t)} | y^{(t)})}{q(y^{(t)} | x^{(t)})}, 1 \right). \quad (4.47)$$

Avant d'adapter l'étape Metropolis à notre problème de ré-échantillonnage des paramètres par un facteur scalaire $c > 0$, nous redéfinissons par souci de simplicité, et de façon strictement équivalente :

$$\theta^{(h)} = \{\beta^{(h)}; S^{(h)}\}, \quad (4.48)$$

avec S , la matrice triangulaire inférieure issue de la décomposition de Cholesky de la matrice Ω . La distribution cible est dans notre modèle bayésien la distribution *a posteriori* des paramètres : $f(x^{(h)}) = P(\theta^{(h)} | X, Y)$. Le rapport des densités *a posteriori* se réduit au simple rapport des densités *a priori* en vertu de la non identification de la fonction de vraisemblance. Puisque $L(Y | X, c\theta^{(h)}) = L(Y | X, \theta^{(h)})$, alors :

$$\frac{f(y^{(h)})}{f(x^{(h)})} = \frac{L(Y | X, c\theta^{(h)})}{L(Y | X, \theta^{(h)})} \times \frac{P(c\theta^{(h)})}{P(\theta^{(h)})} = \frac{P(c\theta^{(h)})}{P(\theta^{(h)})}. \quad (4.49)$$

Compte tenu de l'indépendance des distributions *a priori* :

$$\frac{P(\theta_M^{(h)})}{P(\theta^{(h)})} = \frac{P(c\beta^{(h)}) \times P(cS^{(h)})}{P(\beta^{(h)}) \times P(S^{(h)})}. \quad (4.50)$$

ANDERSON (1984) nous donne les théorèmes permettant de retrouver l'expression de la densité *a priori* de S :

$$P(S) \propto |V|^{-(v/2)} |S|^{-(v+p)} \prod_{j=1}^{p-1} s_{jj}^{p-j} \exp\left(-\frac{1}{2} \text{tr} V^{-1} (SS')^{-1}\right). \quad (4.51)$$

En remplaçant par leur expression, on obtient :

$$\begin{aligned} \frac{P(\theta_M^{(h)})}{P(\theta^{(h)})} &= \frac{\exp\left(-0.5 (c\beta^{(h)} - b)' A^{-1} (c\beta^{(h)} - b)\right)}{\exp\left(-0.5 (\beta^{(h)} - b)' A^{-1} (\beta^{(h)} - b)\right)} \times \dots \\ &\quad \frac{|cS^{(h)}|^{-(v+p)} \prod_{j=1}^{p-1} c(S_{jj}^{(h)})^{p-j} \exp\left(-0.5 \text{tr} \left(V^{-1} (cS^{(h)} S^{(h)'} c)^{-1}\right)\right)}{|S^{(h)}|^{-(v+p)} \prod_{j=1}^{p-1} (S_{jj}^{(h)})^{p-j} \exp\left(-0.5 \text{tr} \left(V^{-1} (S^{(h)} S^{(h)'})^{-1}\right)\right)}, \end{aligned} \quad (4.52)$$

puis en simplifiant, on aboutit à :

$$\frac{P(\theta_M^{(h)})}{P(\theta^{(h)})} = c^{-(p-1)(v+0.5p)} \exp \left(-0.5 \begin{pmatrix} (c^2 - 1) \beta^{(h)'} A^{-1} \beta^{(h)} - 2(c-1) \beta^{(h)'} A^{-1} b + \dots \\ \dots (c^{-2} - 1) \text{tr}(V^{-1} S^{(h)} S^{(h)'}) \end{pmatrix} \right). \quad (4.53)$$

La distribution instrumentale doit autoriser les allers comme les retours dans le ré-échelonnage des paramètres afin de préserver la récurrence de la chaîne des $\theta^{(h)}$: ainsi, un changement d'échelle n'est permis que si son retour à l'état initial est possible :

$$q(\theta^{(h)} | c\theta^{(h)}) > 0 ; q(c\theta^{(h)} | \theta^{(h)}) > 0.$$

Pour effectuer ou non le changement d'échelle, l'étape Metropolis propose un candidat $\theta_M^{(h)}$ échantillonné dans une distribution de répartition $Q(\theta_M^{(h)} | \theta = \theta^{(h)})$, définie pour $\{\theta_M : \theta_M = C\theta^{(h)}\}$, avec C une variable aléatoire continue définie sur \mathbb{R}^{+*} .

Pour échantillonner $\theta_M^{(h)}$, il suffit d'échantillonner un facteur d'échelle lors du h -ième cycle de Gibbs, $c^{(h)}$ dans la distribution de C , univariée, de répartition F_C et de densité f_C :

$$\begin{aligned} Q(\theta_M^{(h)} | \theta^{(h)}) &= \Pr(\theta_M \leq \theta_M^{(h)} | \theta = \theta^{(h)}) \\ &= \Pr(C\theta^{(h)} \leq \theta_M^{(h)}) = \Pr(C \leq \theta_M^{(h)} / \theta^{(h)}) = F_C \left(\frac{\theta_M^{(h)}}{\theta^{(h)}} \right). \end{aligned} \quad (4.54)$$

En dérivant ce résultat par θ_M , nous formulons la densité de la distribution instrumentale. Evaluée en $\theta_M^{(h)}$, nous avons :

$$q(\theta_M^{(h)} | \theta^{(h)}) = \frac{1}{\theta^{(h)}} f_C \left(\frac{\theta_M^{(h)}}{\theta^{(h)}} \right). \quad (4.55)$$

NOBILE (1998) suggère d'utiliser une distribution exponentielle pour C : $C \sim \exp(1)$. L'auteur aboutit finalement à :

$$\frac{q(\theta^{(h)} | \theta_M^{(h)})}{q(\theta_M^{(h)} | \theta^{(h)})} = \frac{\theta^{(h)}}{\theta_M^{(h)}} \exp \left(\frac{\theta_M^{(h)}}{\theta^{(h)}} - \frac{\theta^{(h)}}{\theta_M^{(h)}} \right) = \frac{1}{c^{(h)}} \exp \left(c^{(h)} - \frac{1}{c^{(h)}} \right), \quad (4.56)$$

avec $c^{(h)}$ ici tiré dans une distribution $\exp(1)$. Les formules (4.53) et (4.56) permettent d'évaluer, selon (4.47), la probabilité $\rho(\theta^{(h)}, \theta_M^{(h)})$ de retenir $\theta_M^{(h)}$ au terme du h -ième cycle d'échantillonnage, contre $1 - \rho(\theta^{(h)}, \theta_M^{(h)})$ pour $\theta^{(h)}$.

Utiliser conjointement dans une procédure d'échantillonnage les algorithmes de Gibbs et de Metropolis donne naissance à un algorithme MCMC dit « hybride⁸⁸ ». Quelques détails sur sa convergence sont donnés en annexe dans NOBILE (1998).

4.6 Contrôles de convergence

S'il est pratiquement acquis, dans la grande majorité des cas et sous réserve que les paramètres simulés soient bien identifiables *a posteriori*, que les chaînes de Markov générées par des méthodes MCMC convergent vers une distribution stationnaire⁸⁹, le problème pratique est de déterminer ce moment où la chaîne sort de son étape « burn-in » de convergence. Autrement dit, il est question de déterminer l'itération à partir de laquelle on peut raisonnablement dire que les valeurs suivantes issues de la chaîne sont bien des tirages issus de la loi *a posteriori* visée.

L'examen visuel de l'évolution des paramètres au fur et à mesure des cycles d'échantillonnage⁹⁰ est le premier contrôle pour diagnostiquer une non convergence apparente. Néanmoins, cette méthode « grossière » ne permet que de constater des non convergences fortes des chaînes d'échantillonnage. Par exemple, le logiciel BUGS (*Bayesian Using Gibbs Sampling*⁹¹, SPIEGELHALTER et al., 1994) permet de suivre chronologiquement l'évolution des paramètres pré-sélectionnés sur différents graphiques. Il permet également de faire concourir plusieurs chaînes parallèles superposées ce qui est très pratique pour « appréhender » visuellement la convergence : des chaînes initialisées sur des valeurs dispersées qui viennent se confondre en déviation et en niveau autour d'une moyenne commune est un bon indice que celle-ci a été atteinte.

⁸⁸ Ou « *Metropolis within Gibbs* ».

⁸⁹ « MCMC algorithms will get to the desired posterior density for a wide class of models, even though it makes a long time to get there », JACKMAN (2001).

⁹⁰ Egalement appelé « historique d'échantillonnage ».

⁹¹ www.mrc-bsu.cam.ac.uk/bugs/.

Néanmoins, il est plus rigoureux de faire reposer le diagnostic de convergence sur des critères statistiques plus objectifs que la perception visuelle.

COWLES et CARLIN (1996), BROOKS et ROBERTS (1998) ont proposé un état de l'art des techniques permettant d'évaluer la convergence des chaînes de Markov générées par les algorithmes de Monte Carlo. Deux familles de diagnostics cohabitent. La première regroupe des méthodes d'évaluation fondées sur des techniques *mono-chaînes* : la convergence est appréciée en exploitant une chaîne unique d'échantillonnage. Dans le chapitre précédent, nous avons par exemple utilisé le diagnostic de GEWEKE (1992), qui propose de traiter la séquence de chaque paramètre comme une série temporelle, en utilisant les caractéristiques de sa densité spectrale. En application du théorème de la limite centrale, son diagnostic repose sur l'examen d'une statistique dont la distribution théorique converge vers celle d'une loi normale centrée réduite lorsque la chaîne est stationnaire. Egalement, HEIDELBERGER et WELCH (1983) utilisent la densité spectrale de la chaîne et proposent un diagnostic fondé sur le test d'adéquation d'une fonction avec la distribution d'un pont brownien. Puisque nous ne les avons pas utilisées dans ce chapitre, nous ne développerons pas davantage ces méthodes (parmi d'autres).

La seconde famille de diagnostics rassemble les techniques dites *multi-chaînes*. Deux méthodes couramment employées dans la communauté statistique utilisant les méthodes MCMC sont celles de GELMAN et RUBIN (1992), et de BROOKS et GELMAN (1998). Celles-ci reposent sur la manipulation de plusieurs chaînes, initialisées indépendamment et en des points très dispersés.

4.6.1 Le diagnostic de GELMAN et RUBIN

Le diagnostic de GELMAN et RUBIN (1992) se fonde sur l'appréciation d'un rapport de variance (GELMAN, 1996). Pour chaque paramètre scalaire θ_r^* contenu dans θ^* , on calcule tout d'abord la variance *between* (notée B_r) des moyennes empiriques des m chaînes concourues parallèlement :

$$B_r = \frac{z}{m-1} \sum_{c=1}^m (\bar{\theta}_{rc}^* - \bar{\theta}_r^*)^2, \quad (4.57)$$

sachant :

$$\bar{\theta}_{rc}^* = \frac{1}{z} \sum_{h=1}^z \theta_{rc}^{*(h)} \quad , \quad \bar{\theta}_r^* = \frac{1}{m} \sum_{c=1}^m \bar{\theta}_{rc}^* ,$$

avec $\theta_{rc}^{*(h)}$ le h -ième échantillonnage du r -ième paramètre scalaire de θ^* dans la c -ième chaîne, et avec z le nombre total d'échantillonnages. On construit ensuite la moyenne des variances *within* (notée W_r), calculant la dispersion intra-chaîne moyenne :

$$W_r = \frac{1}{m} \sum_{c=1}^m s_{rc}^2, \quad (4.58)$$

avec :

$$s_{rc}^2 = \frac{1}{z-1} \sum_{h=1}^z \left(\theta_{rc}^{*(h)} - \bar{\theta}_{rc}^* \right)^2.$$

On note que la variance *between* (B_r) comprend un facteur z puisqu'elle est fondée sur la dispersion inter-chaîne des moyennes individuelles sur z tirages. A partir d'une moyenne pondérée des quantités B_r et W_r et pourvu que les tirages soient effectués dans la loi stationnaire visée, on peut construire un estimateur sans biais de la variance de θ_r^* , notée :

$$\hat{\sigma}_{\theta_r^*}^2 = \frac{z-1}{z} W_r + \left(\frac{m-1}{m} \right) \frac{1}{z} B_r. \quad (4.59)$$

L'appréciation de la convergence s'effectue en fonction de la valeur du « *potential scale reduction factor* » (PSRF), noté \hat{R}_r^{GR} , calculé par le rapport de variance :

$$\hat{R}_r^{GR} = \frac{\hat{\sigma}_{\theta_r^*}^2}{W}. \quad (4.60)$$

A mesure que les distributions des m chaînes de Markov convergent vers la loi stationnaire visée, la part de la variance *between* B_r/z dans $\hat{\sigma}_{\theta_r^*}^2$ tend vers 0 et la statistique \hat{R}_r^{GR} tend vers 1 lorsque $z \rightarrow +\infty$: les m chaînes doivent se superposer confortablement sur le graphique de l'historique des simulations de θ_r^* . Dans la pratique, on valide l'hypothèse de la convergence lorsque $\hat{R}_r^{GR} < 1.2$ pour chaque paramètre scalaire θ_r^* de θ^* . Pour des valeurs supérieures, il convient d'augmenter la séquence d'échantillonnage dans l'étape « *burn in* ».

4.6.2 Le diagnostic de BROOKS et GELMAN

BROOKS et GELMAN (1998) ont proposé une alternative, également très simple, fondée sur la longueur des intervalles de crédibilité empiriques pour chaque paramètre scalaire θ_r^* de θ^* . Dans chacune des m chaînes concourues, il s'agit d'identifier dans leur fonction de répartition empirique les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ qui encadrent $(1 - \alpha)\%$ de la distribution des θ_{rc}^* , et d'en déduire la longueur des intervalles par simple différence, notée $dist_{rc}$ pour la c -ième chaîne. En cumulant les chaînes pour n'en former qu'une seule, obtenant ainsi une série de $z \times m$ tirages des θ_r^* , on recherche les quantiles de même ordre dont on déduit similairement la longueur de l'intervalle de crédibilité, notée $dist_{rTOT}$. La statistique \hat{R}_r^{BG} de Brooks-Gelman est définie par le ratio suivant :

$$\hat{R}_r^{BG} = dist_{rTOT} / \left(m^{-1} \sum_{c=1}^m dist_{rc} \right) \quad (4.61)$$

Lorsque l'ensemble des m chaînes a convergé, celles-ci doivent présenter des longueurs d'intervalle individuellement très proches de la longueur d'intervalle de la chaîne cumulée si z est suffisamment grand. Ainsi, la convergence est acquise lorsque les valeurs de \hat{R}_r^{BG} avoisinent 1. Pour tous les éléments θ_r^* de θ^* , on fixe en pratique à 1.2 le seuil des \hat{R}_r^{BG} en dessous duquel la convergence est diagnostiquée. Ce diagnostic est dit *non paramétrique* puisqu'il est uniquement fondé sur l'information donnée par les intervalles de crédibilité.

4.7 Les données « Parc Auto »

L'enquête annuelle Parc Auto est, rappelons le, constituée de deux fichiers de données. Le premier concerne la description désagrégée des ménages. Il détaille individuellement leurs conditions de vie, l'équipement de leur foyer, ainsi que les caractéristiques de leurs membres (âge, sexe, profession...). Chaque ménage possède un numéro d'identifiant unique qui permet un appariement exact avec le second fichier de données, relatif à la description de ses automobiles et de ses utilisateurs. Ce fichier « voitures » renseigne notamment de façon très fine les caractéristiques physiques des automobiles à la disposition des ménages⁹² (niveau de gamme,

⁹² Jusque trois automobiles par ménage.

carburant utilisé, état d'acquisition, kilométrage...). Pour les besoins de notre étude, nous avons sélectionné dans les vagues annuelles d'enquête de 1997 à 2001 de Parc Auto les automobiles acquises en cours d'année par les agents et convenablement renseignées. Les automobiles électriques, hybrides, roulant au GPL ont été écartées : leur type étant trop peu répandu entre 1997 et 2001 (systématiquement moins de 1% des occurrences d'acquisition par vague) pour constituer une véritable modalité de choix pour les consommateurs. Au final, nous disposons de 6633 observations de nouvelles acquisitions pour estimer le modèle MNP, réparties en 1204 observations en 1997 (18.2% de l'échantillon total), 1350 en 1998 (20.4%), 1288 en 1999 (19.4%), 1359 en 2000 (20.5%) et 1432 en 2001 (21.6%).

Nous avons retenu un total de huit modalités de choix pour l'acquisition d'un nouveau modèle automobile, recoupant deux caractéristiques de motorisation (essence ou diesel), deux caractéristiques de gamme (bas et haut), puis deux états possibles à l'achat du véhicule (neuf et occasion). La motorisation du véhicule acquis est renseignée dans l'enquête Parc Auto par le type de carburant principal utilisé pour alimenter le moteur. Sans problème, les voitures catégorisées « essence » sont celles qui utilisent un supercarburant pour rouler (super, super sans plomb 95, super sans plomb 98). Celles-ci représentent 58% des acquisitions dans l'échantillon. Similairement, les voitures « diesel », qui composent 42% de l'effectif total sont celles ayant utilisé le plus souvent du gazole comme carburant.

L'état du véhicule à l'achat est sans équivoque puisqu'il est explicitement demandé dans le questionnaire dont est issue l'enquête, et codé à l'aide d'une variable binaire. Ainsi, nous observons dans nos données que sur 10 véhicules, 6 ont été acquis sur le marché de l'occasion, contre 4 sur le marché du neuf⁹³.

En accord avec l'analyse concurrentielle du marché automobile européen (VERBOVEN, 2002), les échelles de gamme retenues par la commission européenne et par le journal des cotes « L'Argus » sont très proches de la classification adoptée dans l'enquête Parc Auto. Celle-ci décline les véhicules en quatre niveaux explicites : bas (B), moyen inférieur (M1), moyen supérieur (M2), haut de gamme et monospaces (H) ; et distingue les véhicules 4×4 (TT). Afin d'éviter la multiplication des alternatives de choix, nous avons opté pour un découpage plus

⁹³ Parmi leurs résultats sur l'étude des « nouveaux diésélistes », HIVERT et CERRI (2004) montrent qu'au moins sur les années 1994-2000, état à l'achat et motorisation se combinent de façon très différenciée, puisque lors de remplacements, 6 automobilistes sur 10 « fidèles au diesel » achètent leur nouveau véhicule neuf, tandis que 6 « nouveaux diésélistes » (ceux qui passent de l'essence au diesel) ont désormais recours à l'occasion.

simple avec une terminologie un peu différente, séparant « gamme inférieure » (B+M1, 72% des acquisitions) et « gamme supérieure » (M2+H+TT, 28% des acquisitions). Pour alléger la rédaction et faciliter la lecture, adoptons immédiatement la signalétique de la table 4.1 pour désigner les modèles automobiles. La table 4.2 renseigne la répartition des effectifs et décrit leur évolution entre les huit modalités d'acquisition.

TABLE 4.1 : Codes de désignation des alternatives automobiles

Caractéristiques	Code	Véhicule
Gamme	P	Gamme inférieure (Fiat Panda, Opel Corsa, VW Golf, Renault Clio, Seat Ibiza, Peugeot 206,...)
	G	Gamme supérieure (Peugeot 406, Opel Vectra, Audi A6, BMW 520, Renault Espace, Opel Frontera...)
Motorisation	E	Moteur essence (fonctionnant aux supercarburants)
	D	Moteur diesel (fonctionnant au gazole)
Etat à l'achat	O	Occasion (seconde main ou plus)
	N	Neuf (première main)

Exemples : "PDO" désigne une automobile de gamme inférieure, diesel et d'occasion. "PD" désigne une automobile de gamme inférieure et diesel (neuve ou d'occasion). "P" désigne une automobile de gamme inférieure (neuve ou d'occasion, essence ou diesel), etc...

TABLE 4.2 : Effectifs automobiles par année d'observation et par type de modalité d'acquisition

Type/Année	1997	1998	1999	2000	2001	1997-2001
PEO	329 (27.3%)	401 (29.7%)	339 (26.3%)	342 (25.2%)	404 (28.2%)	1815 (27.4%)
PEN	243 (20.2%)	282 (20.9%)	268 (20.8%)	254 (18.7%)	195 (13.6%)	1242 (18.7%)
PDO	146 (12.1%)	168 (12.4%)	189 (14.7%)	206 (15.2%)	220 (15.4%)	929 (14.0%)
PDN	110 (9.1%)	132 (9.8%)	142 (11.0%)	199 (14.6%)	222 (15.5%)	805 (12.1%)
GEO	115 (9.6%)	104 (7.7%)	99 (7.7%)	100 (7.4%)	112 (7.8%)	530 (8.0 %)
GEN	66 (5.5%)	69 (5.1%)	35 (2.7%)	50 (3.7%)	36 (2.5%)	256 (3.9%)
GDO	128 (10.6%)	123 (9.1%)	135 (10.5%)	139 (10.2%)	149 (10.4%)	674 (10.2%)
GDN	67 (5.6%)	71 (5.3%)	81 (6.3%)	69 (5.1%)	94 (6.6%)	382 (5.8%)
Total	1204 (100%)	1350 (100%)	1288 (100%)	1359 (100%)	1432 (100%)	6633 (100%)

Notes : types définis en table 4.1. Source : Parc Auto, vagues 1997-2001.

Les variables explicatives retenues pour modéliser ce choix à l'acquisition tiennent tout d'abord compte de la motorisation passée des individus. En effet, si l'on souhaite révéler l'existence de comportements d'habitude, ou la fidélité des agents pour certains modèles d'automobile, les caractéristiques du véhicule remplacé doivent être considérées. Plus particulièrement à notre cas d'étude, quatre variables indicatrices relatives à l'ancien véhicule ont été créées : celles du type de motorisation (D ou E)⁹⁴, et celles de la gamme (P ou G)⁹⁴. Dans notre échantillon, une acquisition remplace un véhicule essence dans 48% des cas (un véhicule diesel dans 30% des cas). Environ une acquisition sur deux vient en remplacement d'un véhicule bas de gamme, et une sur quatre remplace une automobile de haut de gamme. Pour les agents accédant à leur première voiture, la création d'une indicatrice supplémentaire (pas de véhicule précédent) a été nécessaire. Notamment, ces acquisitions « brutes », c'est-à-dire sans remplacement d'un ancien véhicule, concernent près d'une automobile sur quatre dans notre source de données. Ainsi, l'introduction de ces variables comme éléments explicatifs confère au modèle la possibilité d'établir quelques conclusions intéressantes sur la dynamique comportementale à l'acquisition d'automobile.

De plus, Parc Auto détaille convenablement la catégorie socio-professionnelle (CSP) d'appartenance des utilisateurs principaux des voitures. Ces classes ont la vertu de regrouper des individus assez homogènes, notamment en termes de revenus. Ainsi, notre étude utilise huit indicatrices de CSP, et est en mesure de révéler des singularités entre classes de profession, pour répondre à la question phare de ce chapitre : « qui achète quoi ? ».

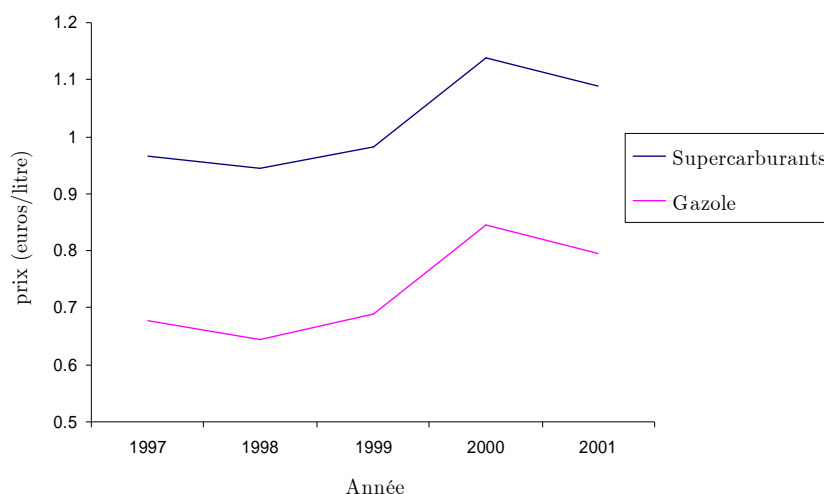
L'analyse descriptive de l'échantillon montre en moyenne que sur cinq automobiles, une est acquise par un agent de profession intermédiaire, une autre par un employé, et une autre par un retraité. Une voiture sur dix est l'acquisition d'un cadre, alors que la part des agriculteurs représente moins de 1% des acquisitions. Les parts respectives des acquisitions faites par les artisans, les ouvriers, et les personnes sans activité professionnelle sont de 2%, 15% et 12%.

L'équation fonctionnelle d'estimation intègre également des caractéristiques socio-démographiques comme facteurs explicatifs pouvant affecter, semble-t-il, le choix du type de véhicule. Notamment, la différenciation du sexe de l'utilisateur pourra mettre en lumière de possibles différences comportementales d'acquisition entre les hommes, qui représentent deux acquéreurs sur trois dans l'échantillon, et les

⁹⁴ Cf. table 4.1.

femmes. Par ailleurs, l'âge de l'utilisateur principal doit être considéré pour notamment neutraliser les changements de préférence des agents au cours du temps. Mais aussi, en reflétant leur position au cours du cycle de vie, l'âge des agents est étroitement corrélé avec d'autres déterminants potentiellement explicatifs, mais parfois inobservés, tels que la structure familiale du ménage, la situation dans l'emploi, le degré d'endettement pour l'habitat... L'âge moyen des acquéreurs de notre source de données est 44 ans. Pour prendre en compte d'éventuelles non linéarités, une variable d'âge mise au carré est également introduite comme régresseur dans la spécification. Enfin, compte tenu des fortes évolutions des prix des carburants (supers et gazole) depuis 1999, il convient de s'interroger à propos de leurs effets sur le choix d'acquisition. Sur la période d'observation (1997-2001), l'évolution en France des prix de ces énergies fossiles a principalement suivi celle du prix du pétrole brut sans être véritablement influencée par des modifications de taxation. Ces trajectoires presque de conserve pourraient rendre le modèle incapable de suffisamment bien distinguer les effets propres d'une variation du prix du gazole et ceux d'une variation du prix des supercarburants sur le choix des agents, tant leurs évolutions parallèles en font des variables extrêmement corrélées (figure 4.2).

FIGURE 4.2 : Evolution du prix des carburants en France (1997-2001)



Source : Calculs INRETS (2003), à partir du Comité Professionnel Du Pétrole (CPDP), et de l'Observatoire De l'Energie.

A cet égard, il semble préférable d'insérer comme facteur explicatif une variable synthétique du coût des carburants, telle que l'indice des prix, librement disponible auprès de l'INSEE. L'idée sous-jacente est de vérifier l'hypothèse selon laquelle les acquéreurs d'automobile porteront davantage leurs préférences sur des véhicules diesel, bénéficiant d'un carburant moins cher et en moyenne moins énergivores, que sur des véhicules essence lorsque le prix des deux carburants principaux évoluent conjointement à la hausse (et réciproquement). En accord avec la figure 4.2, l'indice agrégé des prix des carburants connaît un creux en 1998 et un pic en 2000 : entre ces deux années, les carburants ont vu leur prix fortement augmenter (+24%).

Enfin, les données décrivent le lieu de résidence de chaque acquéreur, permettant au modèle de mesurer son impact sur le type de véhicule choisi. Plus particulièrement, l'ambition est de vérifier l'existence de comportements spécifiques aux Franciliens. Le découpage géographique classique retenu crée quatre indicatrices résidentielles : Paris (département 75), la petite couronne francilienne (départements 92 93 94), la grande couronne francilienne (départements 77 78 91 95), et la Province (départements hors Île-de-France). Dans notre échantillon, 86% des acquisitions sont l'acte de Provinciaux contre respectivement 7.2%, 4.9%, 1.6% pour les Franciliens de grande couronne, de petite couronne, et résidant à Paris.

La table 4.3 présente le codage utilisé pour l'ensemble des variables explicatives sélectionnées, et la table 4.4 en présente les statistiques descriptives dans notre source de données.

TABLE 4.3 : Codage et libellé des variables explicatives du modèle MNP

Variable	Désignation
C	Constante
RX	Indicatrice : 1 si l'agent n'avait pas d'automobile avant d'acquérir son véhicule
EX	Indicatrice : 1 si la précédente automobile était à essence
DX	Indicatrice : 1 si la précédente automobile était diesel
PX	Indicatrice : 1 si la précédente automobile était de gamme inférieure (petite)
GX	Indicatrice : 1 si la précédente automobile était de gamme supérieure (grande)
AGE	Age de l'agent (en années)
AGE2	Age ² /100 de l'agent
HOM	Indicatrice : 1 si l'agent est un homme
FEM	Indicatrice : 1 si l'agent est une femme
CSP1	Indicatrice : 1 si l'agent appartient à la CSP des Agriculteurs
CSP2	Indicatrice : 1 si l'agent appartient à la CSP des Artisans
CSP3	Indicatrice : 1 si l'agent appartient à la CSP des Cadres (ou assimilés)
CSP4	Indicatrice : 1 si l'agent appartient à la CSP des Professions Intermédiaires
CSP5	Indicatrice : 1 si l'agent appartient à la CSP des Employés
CSP6	Indicatrice : 1 si l'agent appartient à la CSP des Ouvriers
CSP7	Indicatrice : 1 si l'agent appartient à la CSP des Retraités
CSP8	Indicatrice : 1 si l'agent appartient à la CSP des « sans activité professionnelle »
PARIS	Indicatrice : 1 si l'agent réside à Paris
PC	Indicatrice : 1 si l'agent réside en petite couronne francilienne
GC	Indicatrice : 1 si l'agent réside en grande couronne francilienne
PROV	Indicatrice : 1 si l'agent réside en Province (hors Île-de-France)
IPC	Indice des prix des carburants

TABLE 4.4 : Statistiques descriptives (annuelles) des variables explicatives

Variable	Moyenne (1997)	Moyenne (1998)	Moyenne (1999)	Moyenne (2000)	Moyenne (2001)	Moyenne (1997-2001)
<i>Indicatrices des caractéristiques des véhicules remplacés</i>						
RX	0.238	0.230	0.234	0.216	0.228	0.229
EX	0.498	0.496	0.464	0.476	0.449	0.476
DX	0.264	0.274	0.303	0.308	0.323	0.295
PX	0.498	0.515	0.484	0.552	0.536	0.518
GX	0.264	0.255	0.282	0.232	0.236	0.253
<i>Age de l'acquéreur</i>						
AGE	43.537	44.046	43.894	43.633	43.087	43.632
AGE2	21.458	22.017	21.706	21.462	20.988	21.519
<i>Indicatrices du genre des acquéreurs</i>						
FEM	0.372	0.354	0.346	0.367	0.384	0.365
HOM	0.628	0.646	0.654	0.633	0.616	0.635
<i>Indicatrices de catégories socio-professionnelles des acquéreurs</i>						
CSP1	0.012	0.007	0.009	0.006	0.003	0.007
CSP2	0.029	0.015	0.018	0.024	0.019	0.021
CSP3	0.081	0.110	0.091	0.090	0.097	0.094
CSP4	0.203	0.174	0.199	0.205	0.207	0.198
CSP5	0.182	0.188	0.195	0.205	0.212	0.197
CSP6	0.148	0.145	0.160	0.152	0.152	0.151
CSP7	0.205	0.227	0.220	0.194	0.191	0.207
CSP8	0.140	0.133	0.109	0.125	0.120	0.125
<i>Indicatrices de localisation résidentielle des acquéreurs</i>						
PARIS	0.015	0.021	0.014	0.015	0.015	0.016
PC	0.049	0.047	0.044	0.057	0.047	0.049
GC	0.066	0.070	0.068	0.070	0.083	0.072
PROV	0.870	0.862	0.874	0.858	0.855	0.864
<i>Indice des prix des carburants (base 100 en 1998)</i>						
IPC	1.032	1.000	1.044	1.237	1.178	1.101
(Effectifs)	(1204)	(1350)	(1288)	(1359)	(1432)	(6633)

Source : parc Auto, vagues 1997-2001.

4.8 Résultats d'estimation

La version bayésienne du modèle MNP, dont la résolution a été détaillée en section 4.4, a été ajustée sur les données de Parc Auto décrites en section 4.7. Pour l'identification, la modalité $p = \text{GDN}$ a été retenue comme référence d'estimation, et la variance de ε_{PEO} a servi pour la normalisation de θ en θ^* (i.e. $\Omega_{11} = \Omega_{PEO,PEO}$). Les paramètres des distributions *a priori* ont été choisis de manière à laisser l'information contenue dans les données dominer très largement les distributions *a posteriori* conditionnelles des paramètres du modèle (section 4.5.2). Concrètement, les valeurs suivantes ont été utilisées pour caractériser des distributions *a priori* (4.25) et (4.26) particulièrement diffuses :

- pour β : $b = 0 \times e_{R \times (p-1)}, A = 100 \times I_{R \times (p-1)},$ (4.62)

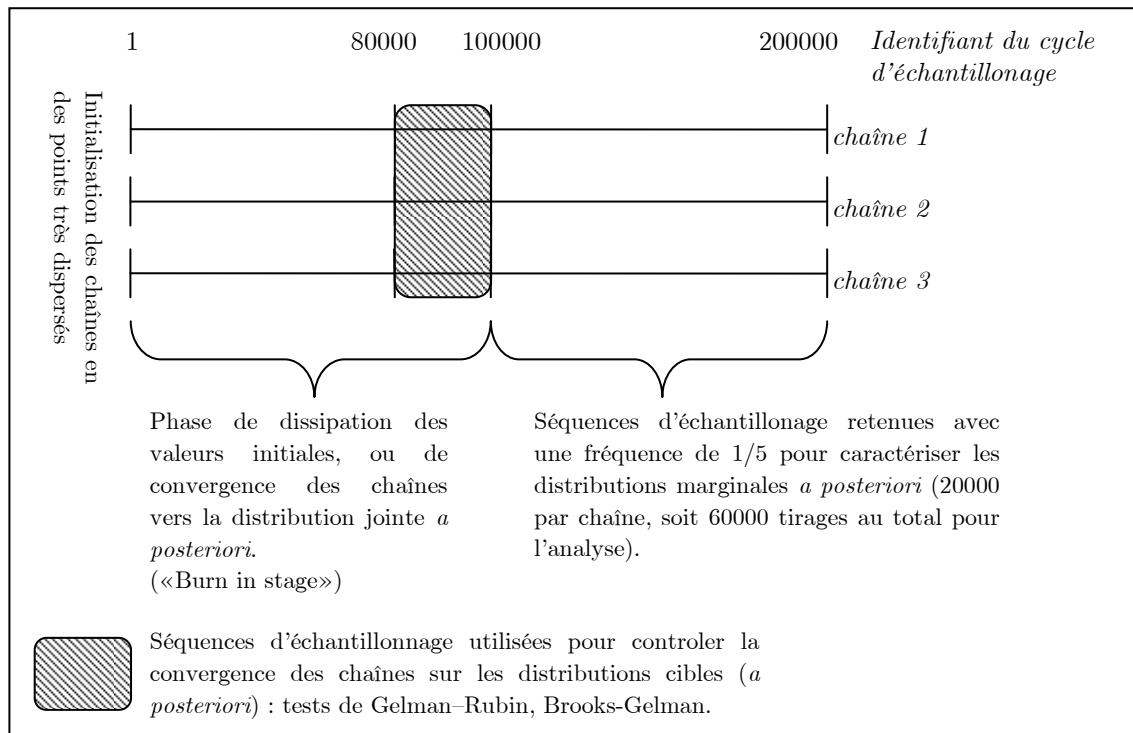
- pour G : $v = 9, V = 9^{-1} \times I_{p-1},$ (4.63)

où $e_{R \times (p-1)}$ est un vecteur colonne de 1 avec $R \times (p-1)$ lignes, R étant le nombre de variables explicatives. Nous avons fait concourir $m = 3$ chaînes d'échantillonnage de $\theta^{(h)} = \{\beta^{(h)}, G^{(h)}\}$ jusque $h = 200000$ itérations, dont sont issues trois chaînes de $\theta^{*(h)} = \{\beta^{*(h)}, \Omega^{*(h)}\}$. Ces chaînes ont été indépendamment initialisées en des points très dispersés, sélectionnés de façon aléatoire par un tirage dans leurs diffuses distributions *a priori*. Pour les itérations h comprises dans $]80000, 100000]$, les distributions empiriques marginales de chaque scalaire contenu dans les matrices β^* et Ω^* sont très voisines entre les trois chaînes. Pour chacun, les diagnostics de Gelman-Rubin et de Brooks-Gelman effectués à partir des trois chaînes entre la 80000-ième et la 100000-ième itération concluent à rejeter l'hypothèse de non convergence vers leur distribution marginale *a posteriori* au terme du 100000-ième cycle d'échantillonnage. En effet, les statistiques \hat{R}_r^{GR} et \hat{R}_r^{BG} correspondantes se sont toutes avérées inférieures à 1.05, et donc largement inférieures au seuil préconisé de 1.2, au delà duquel la non convergence est affichée⁹⁵. Forts de ce résultat, nous pouvons conclure avec confiance que les trois

⁹⁵ Notre modèle compte 154 paramètres composant θ^* : nous avons ainsi calculé 308 statistiques de convergence. Puisqu'elles ont toutes été de bons indices, unanimes pour suggérer que la convergence de chaque paramètre scalaire a été atteinte, nous allégeons notre rédaction d'un grand tableau les reportant. Nous le tenons néanmoins à disposition sur demande.

chaînes d'échantillonnage de $\theta^{*(h)} = \{\beta^{*(h)}, \Omega^{*(h)}\}$ pour $h > 100000$ correspondent à des tirages dans $P(\beta^*, \Omega^* | Y, X)$. Finalement, nous ne conservons des trois chaînes entières que leurs 100000 dernières itérations avec une fréquence de 1/5 (soit 20000 pour chaque chaîne), et l'analyse d'estimation est effectuée à partir de leur cumul, qui correspond à un « output » final de 60000 tirages dans $P(\beta^*, \Omega^* | Y, X)$, noté $\overline{\theta}^{*(l)}$. Une représentation schématique de notre méthodologie est donnée dans la figure 4.3.

FIGURE 4.3 : Traitement multi-chaîne des historiques d'échantillonnage



La table 4.5 reporte les moyennes $\hat{\beta}^*$ de la distribution marginale *a posteriori* de β^* calculées à partir de $\overline{\theta}^{*(l)}$. De la même façon, la table 4.6 reporte les moyennes $\hat{\Omega}^*$ de la distribution marginale *a posteriori* de Ω^* issues de $\overline{\theta}^{*(l)}$. Dans les deux tables, les intervalles de crédibilité à 90% pour chaque paramètre ont été obtenus par détermination des percentiles d'ordre 0.05 et 0.95 de la distribution marginale empirique de leurs simulations depuis $\overline{\theta}^{*(l)}$.

TABLE 4.5 : Moyennes a posteriori des paramètres β^* ($\hat{\beta}^*$)

Variable	PEO	PEN	PDO	PDN	GEO	GEN	GDO
Référence : PX & EX (PEX)							
RX	0.088 [-0.051 ; 0.215]	-0.792 [-1.116 ; -0.534]	-0.098 [-0.266 ; 0.047]	-0.115 [-0.276 ; 0.033]	-0.115 [-0.282 ; 0.029]	-0.289 [-0.474 ; -0.134]	0.150 [-0.023 ; 0.348]
DX	-1.100 [-1.298 ; -0.916]	-0.775 [-1.081 ; -0.534]	-0.016 [-0.172 ; 0.156]	0.419 [0.197 ; 0.727]	-0.671 [-1.001 ; -0.424]	-0.643 [-0.901 ; -0.438]	0.183 [-0.028 ; 0.428]
GX	-0.674 [-0.858 ; -0.480]	-0.780 [-1.086 ; -0.532]	-0.568 [-0.822 ; -0.354]	-0.685 [-0.972 ; -0.445]	0.132 [-0.032 ; 0.310]	0.092 [-0.035 ; 0.239]	0.418 [0.179 ; 0.723]
AGE	-0.049 [-0.069 ; -0.029]	-0.011 [-0.029 ; 0.005]	-0.047 [-0.070 ; -0.027]	-0.009 [-0.026 ; 0.008]	-0.009 [-0.027 ; 0.010]	-0.009 [-0.026 ; 0.008]	0.023 [-0.002 ; 0.052]
AGE2	0.040 [0.019 ; 0.063]	0.027 [0.009 ; 0.048]	0.037 [0.017 ; 0.060]	0.011 [-0.007 ; 0.030]	0.000 [-0.021 ; 0.021]	0.014 [-0.004 ; 0.033]	-0.039 [-0.074 ; -0.010]
Référence : HOM							
FEM	0.333 [0.239 ; 0.430]	0.246 [0.147 ; 0.364]	0.049 [-0.038 ; 0.139]	-0.042 [-0.137 ; 0.040]	-0.094 [-0.203 ; 0.003]	-0.009 [-0.102 ; 0.082]	-0.048 [-0.164 ; 0.057]
Référence : CSP6							
CSP1	0.104 [-0.328 ; 0.548]	-2.529 [-5.438 ; -0.657]	0.414 [0.040 ; 0.848]	-0.263 [-0.790 ; 0.205]	-0.166 [-0.596 ; 0.237]	-1.745 [-4.375 ; -0.176]	-0.092 [-0.572 ; 0.376]
CSP2	-0.249 [-0.516 ; 0.014]	-0.066 [-0.314 ; 0.174]	-0.329 [-0.596 ; -0.098]	-0.094 [-0.317 ; 0.123]	-0.380 [-0.660 ; -0.145]	0.078 [-0.152 ; 0.313]	-0.445 [-0.768 ; -0.174]
CSP3	-0.762 [-0.956 ; -0.576]	0.019 [-0.143 ; 0.204]	-0.559 [-0.798 ; -0.342]	0.064 [-0.095 ; 0.251]	-0.479 [-0.710 ; -0.290]	0.128 [-0.031 ; 0.313]	-0.567 [-0.825 ; -0.342]
CSP4	-0.392 [-0.547 ; -0.244]	0.048 [-0.093 ; 0.210]	-0.389 [-0.570 ; -0.230]	0.137 [0.000 ; 0.310]	-0.445 [-0.653 ; -0.272]	0.105 [-0.041 ; 0.274]	-0.393 [-0.595 ; -0.217]
CSP5	-0.092 [-0.234 ; 0.045]	0.136 [-0.004 ; 0.294]	-0.129 [-0.270 ; -0.003]	0.168 [0.035 ; 0.331]	-0.092 [-0.228 ; 0.032]	0.108 [-0.044 ; 0.277]	-0.222 [-0.403 ; -0.067]
CSP7	-0.310 [-0.516 ; -0.111]	0.055 [-0.116 ; 0.240]	-0.384 [-0.615 ; -0.187]	-0.009 [-0.171 ; 0.163]	-0.260 [-0.468 ; -0.081]	0.076 [-0.098 ; 0.264]	-0.333 [-0.586 ; -0.120]
CSP8	-0.080 [-0.247 ; 0.085]	0.091 [-0.073 ; 0.267]	-0.103 [-0.265 ; 0.047]	0.007 [-0.153 ; 0.170]	-0.116 [-0.284 ; 0.036]	0.068 [-0.124 ; 0.271]	-0.412 [-0.665 ; -0.198]
Référence : PROV							
PARIS	-0.047 [-0.320 ; 0.223]	-0.062 [-0.303 ; 0.172]	-0.344 [-0.680 ; -0.059]	-0.548 [-0.965 ; -0.226]	-0.067 [-0.316 ; 0.170]	-0.014 [-0.229 ; 0.197]	0.108 [-0.185 ; 0.416]
PC	0.087 [-0.088 ; 0.267]	0.282 [0.124 ; 0.467]	-0.109 [-0.303 ; 0.062]	0.007 [-0.150 ; 0.161]	0.220 [0.062 ; 0.406]	0.266 [0.123 ; 0.437]	-0.088 [-0.318 ; 0.116]
GC	0.075 [-0.063 ; 0.214]	0.126 [0.003 ; 0.266]	-0.196 [-0.364 ; -0.056]	-0.044 [-0.169 ; 0.073]	0.020 [-0.104 ; 0.145]	0.117 [0.007 ; 0.239]	-0.084 [-0.250 ; 0.067]
IPC	-0.335 [-0.737 ; 0.064]	-0.745 [-1.215 ; -0.346]	0.253 [-0.100 ; 0.652]	0.649 [0.273 ; 1.127]	-0.234 [-0.626 ; 0.119]	-0.684 [-1.133 ; -0.304]	0.150 [-0.280 ; 0.610]
C	2.147 [1.418 ; 2.897]	1.010 [0.376 ; 1.705]	1.224 [0.606 ; 1.911]	-0.746 [-1.731 ; -0.008]	0.772 [0.136 ; 1.454]	0.617 [0.017 ; 1.250]	-0.814 [-1.955 ; 0.104]

Notes : GDN est pris comme modalité de référence pour l'estimation. Moyennes calculées à partir de $\overrightarrow{\theta^{*(\bar{v})}}$. Intervalles de crédibilité à 90% entre crochets.

TABLE 4.6 : Moyennes a posteriori des paramètres Ω^* ($\hat{\Omega}^*$)

	PEO	PEN	PDO	PDN	GEO	GEN	GDO
PEO	1.000	-0.212 [-0.506 ; 0.073]	0.038 [-0.306 ; 0.359]	-0.081 [-0.442 ; 0.226]	0.096 [-0.194 ; 0.403]	0.202 [-0.010 ; 0.385]	-0.088 [-0.532 ; 0.323]
PEN		0.642 [0.272 ; 1.239]	0.024 [-0.206 ; 0.291]	0.145 [-0.067 ; 0.447]	0.015 [-0.214 ; 0.257]	0.063 [-0.085 ; 0.251]	-0.048 [-0.334 ; 0.231]
PDO			0.614 [0.239 ; 1.169]	-0.035 [-0.272 ; 0.179]	-0.065 [-0.322 ; 0.165]	-0.045 [-0.198 ; 0.102]	-0.132 [-0.627 ; 0.266]
PDN				0.455 [0.175 ; 0.928]	-0.045 [-0.245 ; 0.120]	-0.032 [-0.163 ; 0.081]	0.036 [-0.205 ; 0.335]
GEO					0.477 [0.184 ; 0.985]	0.030 [-0.113 ; 0.181]	0.041 [-0.251 ; 0.345]
GEN		<i>Sym.</i>				0.278 [0.128 ; 0.529]	0.009 [-0.209 ; 0.216]
GDO							0.877 [0.337 ; 1.618]

Notes : GDN est pris comme modalité de référence pour l'estimation. $\text{Var}(\varepsilon_{PEO}) = 1$ pris comme contrainte d'identification. Moyennes calculées à partir de $\overrightarrow{\theta^{*(i)}}$. Intervalles de crédibilité à 90% entre crochets.

La table 4.5 contient les statistiques classiques caractérisant les distributions marginales empiriques des paramètres échantillonnés. Pour chaque composant de $\hat{\beta}^*$, on remarque que l'intervalle de crédibilité à 90% est plutôt symétrique autour de la valeur moyenne. L'analyse graphique individuelle des échantillonnages montre que l'on peut assez confortablement leur ajuster une distribution normale. La table 4.6 contient les résultats d'estimation relatifs à la matrice des covariances des erreurs ε_{ik} : globalement, la méthode d'estimation n'a pas permis de mettre en évidence des covariances significatives⁹⁶. Au premier abord, ce résultat a paru surprenant compte tenu du système de choix. Il semblait effectivement naturel de penser que des modalités étaient suffisamment voisines pour que leurs facteurs inobservés soient significativement corrélés. En pratique, cela nous oblige à examiner plus précisément les propriétés empiriques des erreurs ε_{ik} . Rappelons que :

⁹⁶ En cette circonstance, une prolongation empirique possible à notre travail serait notamment d'estimer le modèle sous contrainte que les covariances de Ω^* sont toutes nulles, puis de le comparer avec celui présenté dans ce chapitre à l'aide du facteur de Bayes.

$$\Omega_{kk'}^* = \Omega_{11}^{-1} \text{cov}(\varepsilon_{ik}, \varepsilon_{ik'}) = \Omega_{11}^{-1} \text{cov}((v_{ik} - v_{ip}), (v_{ik'} - v_{ip})) = \text{var}(v_{i1} - v_{ip})^{-1} (\text{cov}(v_{ik}, v_{ik'}) - \text{cov}(v_{ik}, v_{ip}) - \text{cov}(v_{ik'}, v_{ip}) + \text{var}(v_{ip})). \quad (4.64)$$

Cette décomposition de $\Omega_{kk'}^*$ suggère que les non significativités observées des éléments de covariance $\hat{\Omega}_{kk'}^*$ pourraient résulter simultanément d'une non corrélation entre les aléas v_{ik} et d'une faible variance pour v_{ip} comparativement à v_{i1} , faisant ainsi tendre le résultat de (4.64) vers zéro. L'indépendance des termes v_{ik} revient à examiner la propriété IIA des choix, discutée en section 4.4.1. Selon une procédure décrite dans TRAIN (2002), l'hypothèse IIA a été testée en comparant les paramètres estimés de deux modèles logit multinomiaux, le premier considère l'ensemble des huit modalités d'acquisition, alors que le second exclut une alternative de l'espace des choix. Quelle que soit l'alternative exclue, les tests comparatifs de Hausman⁹⁷ acceptent l'hypothèse d'égalité des deux ensembles de paramètres obtenus, validant la propriété IIA du modèle, et concluant ainsi à l'indépendance des erreurs v_{ik} .

On fait également apparaître la variance de v_{ip} en décomposant Ω_{kk}^* :

$$\Omega_{kk}^* = \Omega_{11}^{-1} \text{var}(\varepsilon_{ik}) = \Omega_{11}^{-1} \text{var}(v_{ik} - v_{ip}) = \Omega_{11}^{-1} (\text{var}(v_{ik}) - 2 \text{cov}(v_{ik}, v_{ip}) + \text{var}(v_{ip})). \quad (4.65)$$

Sous la condition IIA précédemment validée, il vient que $\text{cov}(v_{ik}, v_{ip}) = 0$, et donc :

$$\Omega_{kk}^* = (\text{var}(v_{i1}) + \text{var}(v_{ip}))^{-1} (\text{var}(v_{ik}) + \text{var}(v_{ip})). \quad (4.66)$$

Une estimation $\hat{\Omega}_{kk}^*$ de Ω_{kk}^* faible est un indicateur empirique suggérant que $\text{var}(v_{ip})$ est petit relativement à $\text{var}(v_{i1})$, puisque $\text{var}(\cdot)$ est strictement positif. Dans la table 4.6, $\hat{\Omega}_{GEN,GEN}^* = 0.278$ laisse ainsi penser que $\text{var}(v_{ip})$ est effectivement petit. Ce raisonnement joint à la propriété IIA validée par

⁹⁷ Par exemple, sous l'hypothèse H0 d'égalité des paramètres entre les deux modèles MNL, la statistique d'Hausman, qui suit théoriquement une distribution $\chi^2(108)$, vaut 85.66 lorsque l'alternative PDO est retirée du système de choix. La valeur tabulée correspondant au risque d'erreur de première espèce à 5% étant 133.26, l'hypothèse H0 est donc validée par le test. Plus généralement, cette conclusion s'est imposée quelle que soit la modalité retirée de l'ensemble de choix.

l'ajustement de modèles MNL nous semble constituer une explication de la non significativité des Ω_{kk}^* , reportée dans la table 4.6.

La propriété IIA mérite d'être commentée : elle révèle ici que les choix individuels dans notre échantillon n'obéissent à aucun emboîtement décisionnel très précis. Dès lors, si un processus de choix par étapes peut exister, celui-ci doit être variable selon les individus. Par exemple, des acquéreurs peuvent d'abord choisir entre un véhicule E ou D avant de déterminer le niveau de gamme, certains peuvent privilégier le marché d'acquisition (O ou N) avant autre chose, alors que d'autres agents peuvent fonder leur choix en examinant toutes les modalités simultanément, etc...

Selon LUCE (1959), une modélisation qui exhiberait la propriété IIA reposerait sur une spécification appropriée des probabilités de choix, qui inclut des facteurs explicatifs en mesure suffisante d'« absorber » la corrélation entre aléas. Sans être systématique, c'est plus favorablement le cas des variables spécifiques aux alternatives (i.e. associées à des coefficients qui diffèrent pour chaque alternative), dont l'usage est recommandé pour éviter les risques de violation de la propriété IIA (BEN-AKIVA et LERMAN⁹⁸, 1995). Puisque les variables explicatives utilisées dans notre étude sont toutes de cette nature, alors la propriété IIA ne doit, au final, pas sembler si surprenante.

4.9 Exploitation des résultats

Si l'on peut effectivement extraire à partir des résultats d'estimation une mesure concernant l'impact « *ceteris paribus* » d'un changement de caractéristiques individuelles sur le niveau des utilités relatives W^* d'un agent avec les méthodes usuelles, cela ne présente qu'un intérêt particulièrement limité. En effet, une telle démarche ne permet pas de mesurer directement l'impact induit sur la répartition des probabilités de choix entre alternatives, qui dépend à la fois du niveau des espérances des variables latentes normalisées W^* ($E(W_i^*) = X_i\beta^*$) et de la structure générale donnée par le modèle MNP à la matrice des covariances des termes d'erreur Ω^* .

Aux points moyens d'échantillonnages $\hat{\beta}^*$ et $\hat{\Omega}^*$ (tables 4.5 et 4.6), la mesure du système de probabilités estimé d'un agent i réclame l'évaluation d'une

⁹⁸ Cités dans CHOO et MOKHTARIAN (2002).

intégrale multiple dès lors que le nombre d'alternatives est supérieur ou égal à trois. Sa résolution analytique étant très décourageante (voire impossible), l'usage des techniques de simulation est le recours le plus judicieux pour convenablement l'approximer. Parmi les simulateurs de probabilités les plus populaires dans la communauté statistique, on peut citer le simulateur de fréquences brut CFS⁹⁹ (LERMAN et MANSKI, 1981), le simulateur de Stern (STERN, 1992), et le simulateur GHK¹⁰⁰ (HAJIVASSILIOU et al., 1996). On pourra lire ARIAS et COX (1999) pour une introduction et une discussion de ces trois outils.

4.9.1 Le simulateur de fréquence brute CFS

Concernant notre étude, nous avons choisi d'utiliser le simulateur CFS, qui possède de bonnes propriétés usuelles : celui-ci est sans biais, de variance minimale (LERMAN et MANSKI, 1981). Son principe de fonctionnement adapté à notre problème est intuitivement simple : le simulateur effectue T tirages dans la distribution des utilités relatives évaluée aux résultats d'estimation $\hat{\beta}^*$ et $\hat{\Omega}^*$, puis la probabilité estimée du choix d'une alternative j (pour $j \neq p$) est la fréquence de tirages avec laquelle l'utilité relative associée est à la fois positive et maximale sur l'ensemble des utilités relatives. La probabilité estimée du choix de la modalité de référence p correspond à la fréquence de tirages pour laquelle toutes les variables latentes échantillonnées sont simultanément négatives. Schématiquement, les étapes de simulation des probabilités de choix sont les suivantes :

1. Soit Γ la transformation de Cholesky de $\hat{\Omega}^*$, telle que $\Gamma\Gamma' = \hat{\Omega}^*$,
2. Soit A_s le s -ième tirage dans une $\mathcal{N}_{p-1}(0, I_{p-1})$, avec $s \in [1 : T]$,
3. Construire $\hat{W}_{is}^* = X_i\hat{\beta}^* + \Gamma A_s$,
4. Calculer :

$$\widehat{\text{Pr}}(D_i = j) = \frac{1}{T} \sum_{s=1}^T \mathbf{1}(\hat{W}_{ijs} = \max_k(\hat{W}_{iks})) \mathbf{1}(\hat{W}_{ijs} > 0) \text{ pour } j \neq p,$$

$$\widehat{\text{Pr}}(D_i = p) = 1 - \sum_{j=1}^{p-1} \widehat{\text{Pr}}(D_i = j),$$

⁹⁹ CFS pour « Crude Frequency Simulator ».

¹⁰⁰ GHK pour Geweke-Hajivassiliou-Keane, du nom des trois auteurs ayant contribué à l'élaboration du simulateur lisse récursif SRC (« Smooth Recursive Conditioning Simulator ») pour la loi normale multi-dimensionnelle.

où $\mathbf{1}(\cdot)$ est la fonction indicatrice qui vaut 1 si l'argument entre parenthèses est vrai, 0 sinon. Outre l'interprétation classique en termes de probabilités, les mesures $\widehat{\text{Pr}}$ peuvent de façon équivalente être décrites comme des « parts de marché ». Par exemple, si $\widehat{\text{Pr}}(D_i = j)$ mesure la probabilité estimée que l'individu i choisisse l'alternative j , c'est également dans ce contexte la part de marché estimée associée au modèle automobile j pour une « clientèle » d'acquéreurs ayant les caractéristiques de l'individu i . Rappelons que la distribution des probabilités entre les différentes alternatives de choix dépend du niveau de l'espérance des variables latentes, et donc des caractéristiques individuelles. De fait, l'analyse « *ceteris paribus* » des variations du système de probabilités estimées résultant d'un changement de caractéristiques ne peut être considérée qu'au cas par cas. Pour l'appliquer, il convient de déterminer un individu particulier, choisi conventionnellement à la moyenne des caractéristiques de l'échantillon. C'est donc sur la base des caractéristiques de l'individu moyen, parfois appelé individu « synthétique », que s'effectuera l'analyse « toutes autres caractéristiques égales avec l'agent moyen ».

Pour chaque individu examiné dans les tables suivantes (4.7 jusque 4.11), ses probabilités d'acquisition reportées utilisent les moyennes des paramètres échantillonnés données en tables 4.5 et 4.6, et le simulateur CFS avec $s = 25000$ simulations. Usuellement, les écarts-types ou les intervalles de confiance associés sont dérivés de la précision des estimations. Dans notre cas, celle-ci repose sur la variabilité des couples d'échantillonnage $\theta^{*(l)} = \{\beta^{*(l)}, \Omega^{*(l)}\}$ tirés dans la distribution *a posteriori* $P(\beta^*, \Omega^* | Y, X)$. Pratiquement, nous avons sélectionné aléatoirement 1500 couples $\theta^{*(m)} = \{\beta^{*(m)}, \Omega^{*(m)}\}$ (notés $\overrightarrow{\theta^{*(m)}}$) dans l'historique d'échantillonnage $\overrightarrow{\theta^{*(l)}}$ parmi les 60000 à disposition (figure 4.3). Pour chacun, le simulateur CFS a utilisé 2000 simulations¹⁰¹ afin de calculer le vecteur des probabilités d'acquisition $\widehat{\text{Pr}}^{(m)}$ de l'individu considéré. Au final, nous disposons de 1500 vecteurs de ses probabilités d'acquisition permettant de déterminer pour chacune un intervalle de confiance à 95% (reportés entre crochets¹⁰²), par

¹⁰¹ Pour des raisons pratiques liées au temps de calcul, il était impensable d'effectuer 25000 simulations du simulateur CFS pour chacun des 60000 échantillonnages contenus dans $\overrightarrow{\theta^{*(l)}}$. Par ailleurs, l'usage de $s = 2000$ relativement à $s = 25000$ ne semble pas être pénalisant puisque ces deux valeurs donnent lieu à des intervalles de confiance quasi identiques pour les quelques probabilités testées.

¹⁰² On observera que les probabilités reportées dans les tables 4.7-4.11 ne correspondent pas systématiquement au centre de l'intervalle de confiance. Une raison tient au fait que le modèle MNP est non linéaire.

détermination des percentiles d'ordre 0.025 et 0.975. Puisqu'elle utilise plusieurs tirages issus d'une distribution pour caractériser des quantités d'intérêts, notre méthode de détermination des intervalles de confiance est une technique de « *bootstrap paramétrique* ».

Pour comparer les probabilités d'un même système, celui de l'agent i , un outil d'analyse particulièrement intuitif et largement utilisé dans la littérature des modèles de choix discrets est le « *odds* ». Celui-ci (noté O_{A-B}^i) définit le rapport de probabilités entre deux ensembles (A et B) d'alternatives mutuellement exclusives, et s'interprète comme le rapport de « chances » de choisir l'un plutôt que l'autre : $O_{A-B}^i = \Pr(D_i = A) / \Pr(D_i = B)$. Pour comparer les « *odds* » du choix d'un ensemble A d'alternatives sur un autre B entre deux individus i et i' différents, nous adoptons une interprétation en « *odds ratio* » (noté $OR_{A-B}^{i-i'}$, défini par : $OR_{A-B}^{i-i'} = O_{A-B}^i / O_{A-B}^{i'}$).

4.9.2 Effet de la motorisation passée

Une propriété de la spécification est de pouvoir mettre en lumière un effet de dépendance du choix d'acquisition aux caractéristiques du véhicule précédent. L'estimation (table 4.5) montre une significative fidélité des agents au type de véhicule remplacé, tant au niveau de la motorisation (E ou D) que sur le plan du niveau de gamme (P ou G).

Toutes autres caractéristiques égales, la probabilité d'un individu de porter son choix sur une automobile diesel augmente significativement si sa précédente voiture avait également un moteur diesel plutôt qu'essence. Similairement, la probabilité de choisir un véhicule de gamme supérieure augmente significativement lorsque la voiture précédente était elle aussi une voiture haut de gamme, plutôt que de gamme inférieure.

Par ailleurs, les résultats montrent que l'acquisition d'une première voiture par un agent est statistiquement moins examinée sur le secteur des automobiles essence neuves que s'il devait notamment remplacer une automobile PE.

Toutes autres caractéristiques équivalentes avec celles de l'individu synthétique, un agent non-équipé avant l'acquisition possède près de deux fois plus de chances de se porter son choix sur un véhicule essence que sur une voiture diesel : $\hat{O}_{E-D}^{RX} = 1.9$ ($\hat{O}_{E-D}^{PEX} = 3.1$, $\hat{O}_{E-D}^{GEX} = 2.1$). Au contraire, si la voiture précédente de l'agent avait été PD, alors le modèle estime qu'il a près de trois fois

plus de chances de choisir un véhicule diesel qu'essence : $\widehat{O}_{D-E}^{PDX} = 2.9$ ($\widehat{O}_{D-E}^{GDX} = 4$). Ainsi, l'ancien propriétaire d'un véhicule PD a près de neuf fois plus de chances qu'un agent précédemment équipé d'un véhicule PE de choisir un véhicule diesel plutôt qu'essence : $\widehat{OR}_{D-E}^{PDX-PEX} = 9.1$. Similairement, l'agent dont l'ancien véhicule était GD a huit fois plus de chances qu'un agent se séparant d'une voiture GE de choisir un véhicule diesel plutôt qu'essence : $\widehat{OR}_{D-E}^{GDX-GEX} = 8.1$.

TABLE 4.7 : Répartition des probabilités d'acquisition selon le type de véhicule précédent

Type acquis	Type de véhicule précédent(*)				
	RX	PEX	GEX	PDX	GDX
PEO	45.1% [42.8 – 51.0]	34.2% [31.8 – 38.9]	19.7% [16.4 – 24.1]	10.9% [8.0 – 13.1]	4.7% [2.8 – 6.2]
PEN	10.3% [8.0 – 13.2]	32.1% [29.3 – 36.7]	13.9% [10.8 – 16.7]	10.7% [8.0 – 13.5]	3.3% [1.9 – 4.5]
PDO	15.2% [11.6 – 17.8]	12.5% [9.8 – 14.7]	8.3% [5.7 – 10.4]	23.3% [19.8 – 27.3]	13.8% [10.4 – 16.6]
PDN	9.1% [6.5 – 11.3]	6.7% [4.4 – 7.9]	2.7% [1.1 – 3.5]	35.8% [32.9 – 41.2]	15.1% [11.6 – 18.0]
GEO	8.5% [6.0 – 10.7]	7.5% [5.3 – 9.0]	22.2% [18.6 – 27.0]	3.3% [1.6 – 4.2]	8.6% [5.8 – 10.7]
GEN	1.2% [0.3 – 2.0]	2.1% [0.9 – 3.0]	11.5% [8.8 – 14.9]	0.9% [0.2 – 1.4]	3.6% [1.9 – 4.9]
GDO	7.7% [5.2 – 9.5]	4.0% [2.3 – 5.0]	16.0% [12.3 – 19.2]	10.1% [7.2 – 12.2]	30.8% [26.6 – 35.2]
GDN	2.9% [1.6 – 4.3]	0.9% [0.2 – 1.4]	5.7% [3.9 – 8.2]	5.0% [3.3 – 7.2]	20.1% [18.3 – 26.1]

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

Types définis en table 4.1. Pour alléger les écritures : PEX équivaut à l'intersection de PX et de EX (une voiture précédente essence et de gamme inférieure), GEX équivaut à l'intersection de GX et de EX (une voiture précédente essence et de gamme supérieure), etc...

Lorsque l'individu synthétique n'avait pas de véhicule précédent, la part de marché la plus élevée à l'acquisition d'un véhicule est celle des voitures PE (55.4%) et plus généralement celle des voitures de gamme inférieure (79.7%). Dans des mesures comparables, lorsque l'agent moyen remplace son ancien véhicule PE, c'est

pour une automobile de même gamme et utilisant le même carburant avec une probabilité estimée de 66.3%, alors que sa probabilité de choisir plus généralement un véhicule bas de gamme est de 85.5% : l'agent a donc près de six fois plus de chances de choisir un véhicule de gamme inférieure que supérieure. Le constat est beaucoup moins tranché pour l'agent ancien propriétaire d'un GE : en effet, celui-ci a quasiment autant de chances d'acquérir un véhicule PE qu'un nouveau GE. Au contraire, si l'individu synthétique devait choisir de remplacer un véhicule GD, il choisirait dans 51% des cas une automobile similaire. Egalement, la part de marché des véhicules PD pour l'agent détenant auparavant une voiture de même type est de 59.1%. Plus généralement, il choisira près de 8 fois sur 10 un véhicule de bas de gamme.

4.9.3 Effet du sexe de l'acquéreur

Le modèle met également en évidence un comportement d'acquisition significativement différent selon de genre de l'acquéreur. Plus particulièrement, les véhicules PE sont la modalité privilégiée des femmes : à toutes autres choses égales avec l'individu synthétique, la part de marché des véhicules PE est de 53.8% pour une femme, contre 38.5% pour un homme.

TABLE 4.8 : Répartition des probabilités d'acquisition selon le genre

Sexe de l'acquéreur			Sexe de l'acquéreur		
Type acquis	Femme	Homme	Type acquis	Femme	Homme
PEO	32.6% [29.7 – 37.6]	23.0% [20.5 – 26.9]	GEO	7.1% [4.9 – 8.9]	11.5% [9.0 – 13.9]
PEN	21.2% [18.6 – 25.6]	15.3% [12.7 – 18.2]	GEN	2.0% [0.8 – 3.0]	3.5% [2.0 – 4.8]
PDO	15.7% [12.1 – 18.3]	17.3% [14.5 – 20.0]	GDO	8.6% [6.1 – 10.6]	11.1% [8.4 – 12.9]
PDN	9.3% [6.9 – 11.6]	13.2% [10.7 – 16.0]	GDN	3.4% [2.0 – 5.0]	5.0% [3.5 – 7.0]

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

Types définis en table 4.1.

De plus la probabilité d'acquérir une automobile de gamme inférieure pour un homme est de 10 points inférieure à celle d'une femme (78.8% contre 68.8% pour un homme).

Par ailleurs, on observe le même écart de points entre les parts de marché des véhicules essence entre les deux sexe (63% pour une femme et 53.3% pour un homme), une femme présentant un $\hat{O}_{E-D}^{FEM} = 1.7$, contre $\hat{O}_{E-D}^{HOM} = 1.1$ pour un homme. Une femme a ainsi 1.5 fois plus de chances qu'un homme de choisir une automobile essence plutôt que diesel lors de l'acquisition d'une automobile.

Si l'on vient de vérifier la préférence relative des hommes pour les automobiles haut de gamme puis diesel, les résultats montrent que les deux sexes ont approximativement autant de chances de choisir un véhicule neuf ou un véhicule d'occasion : la part de marché estimée des voitures d'occasion est de 64% pour une femme et de 63% pour un homme.

4.9.4 Effet de la catégorie socio-professionnelle (CSP)

L'un des comportements les plus atypiques dans l'acquisition d'une automobile reflété par les résultats du modèle est celui des agents issus de la catégorie des agriculteurs (CSP1). Aux caractéristiques de l'individu moyen, le modèle estime qu'un agriculteur aura recours en moyenne près de 9 fois sur 10 au marché de l'occasion, et choisira un véhicule diesel près de 6 fois sur 10. Plus spécifiquement, sa part de marché associée aux véhicules DO est la plus forte au regard des autres CSP (51.7%). Cependant, il convient d'être prudent avec ces conclusions tant les observations appartenant à cette classe sont peu représentées dans la base de données : celles-ci représentent moins de 1% des effectifs totaux (46 observations sur 6633). Il existe donc un risque plus élevé que dans l'ensemble des autres CSP d'avoir un échantillon insuffisamment représentatif du véritable comportement d'acquisition des agents agriculteurs.

Les résultats d'estimation montrent également que les artisans (CSP2), les cadres (CSP3), les ressortissants des professions intermédiaires (CSP4) et les retraités (CSP7) se distinguent, à toute autre caractéristique égale, des ouvriers (CSP6) en ayant un recours significativement moins fréquent au marché de l'occasion pour s'équiper d'une automobile, quelque soit le niveau de gamme ou le type de motorisation. Sans équivoque, l'effet de revenu est manifestement l'une des origines de cette divergence de comportement, les ouvriers étant comparativement

moins riches. Ce résultat est d'autant plus effectif que les paramètres estimés associés aux modalités des véhicules d'occasion pour les cadres (CSP3), qui constituent la catégorie professionnelle la plus riche de la population, sont les plus négatifs.

TABLE 4.9 : Répartition des probabilités d'acquisition selon la CSP

Type acquis	Catégorie Socioprofessionnelle de l'agent							
	CSP1	CSP2	CSP3	CSP4	CSP5	CSP6	CSP7	CSP8
PEO	31.6% [19.0 – 46.5]	28.6% [21.4 – 39.2]	15.9% [12.1 – 19.9]	23.7% [20.6 – 28.3]	28.3% [24.9 – 33.4]	29.9% [26.1 – 35.5]	26.0% [21.1 – 32.2]	29.9% [25.9 – 35.8]
PEN	1.0% [0.0 – 3.3]	16.3% [10.1 – 25.2]	21.4% [17.4 – 26.4]	19.5% [16.1 – 23.4]	17.3% [13.9 – 21.3]	12.7% [9.5 – 16.3]	19.4% [15.4 – 25.1]	17.5% [13.5 – 22.1]
PDO	38.9% [25.1 – 55.9]	16.9% [9.7 – 23.6]	12.8% [8.8 – 16.0]	14.4% [10.9 – 17.2]	17.2% [13.8 – 20.9]	20.5% [16.6 – 25.0]	13.9% [9.5 – 18.3]	19.5% [15.0 – 23.7]
PDN	4.5% [0.4 – 12.5]	9.9% [4.8 – 15.8]	16.8% [12.5 – 21.4]	15.8% [12.6 – 19.4]	12.1% [9.0 – 15.4]	7.5% [5.0 – 10.0]	10.9% [7.2 – 14.9]	9.1% [5.8 – 11.9]
GEO	8.3% [2.2 – 18.4]	8.1% [3.5 – 13.2]	7.7% [4.8 – 10.5]	6.4% [4.1 – 8.4]	10.5% [7.7 – 13.5]	12.3% [8.7 – 15.4]	10.0% [6.3 – 13.9]	11.0% [7.8 – 14.7]
GEN	1.0% [0.0 – 1.0]	3.9% [1.1 – 8.5]	7.4% [4.6 – 10.6]	4.6% [2.5 – 6.3]	2.3% [0.9 – 3.5]	1.2% [0.3 – 2.2]	3.6% [1.6 – 5.7]	2.3% [0.7 – 4.3]
GDO	11.8% [3.8 – 23.0]	9.1% [4.4 – 14.5]	8.0% [5.0 – 10.7]	9.6% [6.6 – 12.0]	10.0% [6.8 – 12.4]	13.7% [10.2 – 17.1]	10.6% [6.9 – 14.2]	7.4% [4.5 – 10.1]
GDN	2.9% [0.2 – 9.2]	7.2% [3.1 – 13]	9.9% [7.6 – 14.6]	6.0% [4.1 – 8.8]	2.2% [1.1 – 3.9]	2.1% [0.8 – 3.5]	5.5% [3.4 – 9.5]	3.4% [1.4 – 5.7]

Notes : évaluations aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

Types définis en table 4.1, catégories socio-professionnelles définies en table 4.3.

Ainsi, un cadre présente la propension la plus élevée à l'acquisition d'un modèle automobile neuf : aux caractéristiques de l'agent synthétique, celui-ci n'a recours au marché de l'occasion qu'avec une probabilité de 44.5% selon les résultats, alors qu'elle est systématiquement supérieure à 50% dans l'ensemble des autres catégories professionnelles. Notamment pour cet individu, la part de marché des véhicules DN est de loin la plus élevée (26.7%). Relativement aux autres catégories de profession, sa préférence pour les modèles haut de gamme est tout aussi remarquable. A titre de comparaison, cet agent a *ceteris paribus* près de cinq fois

plus de chances qu'un employé (CSP5) de s'équiper d'un véhicule GDN plutôt que d'une autre voiture ($\widehat{OR}_{GDN-\overline{GDN}}^{CSP3-CSP5} = 4.88$)¹⁰³.

Dans une mesure moins tranchée que pour un cadre, mais de façon comparable, un individu ressortissant de la classe des professions intermédiaires (CSP4) présente une propension relativement élevée à recourir au marché des automobiles neuves, la part de marché associée étant de 45.8%. Pour comparaison et aux caractéristiques moyennes, l'agent issu de la CSP4 a environ trois fois plus de chances qu'un ouvrier de choisir un véhicule neuf plutôt que d'occasion ($\widehat{OR}_{N-O}^{CSP4-CSP6} = 2.75$). Hormis le cas spécifique de l'agriculteur, les agents cadres ou de profession intermédiaire présentent, aux caractéristiques de l'individu moyen, les plus fortes propensions à l'achat de véhicules diesel : respectivement 47.5% et 45.9%, alors que cette part de marché estimée est d'environ 6 points inférieure pour les agents employés ou inactifs.

Le comportement d'acquisition d'un ouvrier (CSP6) présente la singularité d'un fort recours au marché de l'occasion : toutes autres caractéristiques égales avec l'agent moyen, celui-ci a près de trois fois plus de chances d'acquérir une automobile de seconde main qu'un véhicule neuf ($\widehat{O}_{O-N}^{CSP6} = 3.24$). Pour cet agent, les parts de marché des véhicules PO (50.4%) et GO (26%) sont les plus élevées, comparativement à celles des autres CSP (CSP1 exceptée). Outre l'explication par le niveau de revenu, cette préférence révélée des ouvriers pour les voitures de seconde main pourrait résulter en partie de leur capacité à pouvoir apprécier la véritable qualité des automobiles proposées sur le marché de l'occasion, ainsi que de leur aptitude à pouvoir réparer eux-mêmes des automobiles plus âgées avec peu d'électronique embarquée.

La proximité des caractéristiques (notamment de revenus) des individus employés (CSP5) et des inactifs (CSP8) avec celles des ouvriers conduit à une majorité de non significativités des coefficients estimés associés aux variables CSP5 et CSP8 dans la table 4.5. Précisément, employés et inactifs présentent des niveaux d'utilité relatives non statistiquement différents de ceux des ouvriers sur l'ensemble des modalités de choix, excepté pour celle des véhicules GDO.

Bien qu'il soit inférieur à celui des ouvriers, on observe un recours élevé aux véhicules d'occasion chez les agents issus de la catégorie des employés et chez les inactifs : aux caractéristiques de l'agent synthétique, les probabilités respectives d'acquérir ces véhicules sont évaluées à 66% et 67.8%. *Ceteris paribus*, un agent

¹⁰³ \overline{GDN} pour l'ensemble des alternatives, exceptée la modalité *GDN*.

appartenant à l'une des catégories précédentes aura approximativement 2.5 fois moins de chances d'acquérir une automobile neuve qu'un individu cadre ($\widehat{OR}_{O-N}^{CSP5-CSP3} = 2.42$; $\widehat{OR}_{O-N}^{CSP8-CSP3} = 2.62$).

4.9.5 Effet de la zone résidentielle

Les résultats d'estimation montrent que l'utilité retirée par l'acquisition d'un véhicule PD est, toutes choses égales par ailleurs, statistiquement moins élevée pour un Parisien que pour un résidant provincial. De même, un Francilien de grande couronne présente une préférence moins élevée qu'un Provincial pour l'acquisition d'une voiture PDO. Les Franciliens de petite couronne se démarquent des résidents provinciaux, en révélant un attrait significativement plus élevé pour les automobiles de type GE.

Aux caractéristiques de l'agent moyen, un Parisien affiche une préférence plus marquée pour les véhicules de gamme supérieure que l'ensemble des autres résidents franciliens, ou que les Provinciaux. En effet, sa probabilité estimée de choisir une voiture haut de gamme est de 37%, alors qu'elle est globalement inférieure de 10 points pour les autres individus. Ainsi, un agent résidant en grande couronne francilienne a 1.6 fois plus de chances qu'un Parisien de choisir un véhicule bas de gamme plutôt que de gamme supérieure ($\widehat{OR}_{P-G}^{GC-Paris} = 1.60$). Relativement à un Provincial et quelle que soit sa zone de résidence, un Francilien présente une préférence singulièrement plus élevée pour les véhicules à moteur essence. Selon les résultats du modèle, le résidant de petite couronne a plus de 2 fois plus de chances d'acquérir un véhicule essence qu'une automobile diesel, contre 1.2 pour l'individu résidant en Province ($\widehat{O}_{E-D}^{Paris} = 1.61$, $\widehat{O}_{E-D}^{GC} = 1.79$). On observe par ailleurs un comportement tout à fait remarquable de l'agent résidant à Paris : son recours aux automobiles diesel s'effectue dans une large mesure plus fréquemment vers les voitures de gamme supérieure au détriment des voitures bas de gamme, alors qu'on observe un comportement opposé chez les résidents des autres zones. Aux caractéristiques moyennes, les probabilités d'acquisition des véhicules GD et PD sont respectivement évaluées à 23.3% et 15.1% pour l'agent parisien, alors qu'elles sont respectivement de 9.9% et 21.9% pour le résidant de petite couronne par exemple. Plus explicitement encore, l'individu parisien a *ceteris paribus* 2.6 fois plus de chances que l'agent résidant en grande couronne francilienne de choisir une voiture GD plutôt qu'une voiture PD ($\widehat{O}_{GD-PD}^{Paris-GC} = 2.61$). Enfin, l'agent de petite

couronne présente la propension la moins élevée à recourir au marché de l'occasion pour l'acquisition d'une automobile ($\widehat{\Pr}(D_{PC} = O) = 57.5\%$).

TABLE 4.10 : Répartition des probabilités d'acquisition selon la zone de résidence

Type acquis	Zone de résidence de l'agent			
	PARIS	PC	GC	PROV
PEO	28.7% [19.4 – 39.6]	25.4% [20.9 – 32.6]	29.0% [24.4 – 35.4]	25.8% [23.6 – 30.1]
PEN	19.2% [11.7 – 29.1]	24.5% [18.8 – 30.5]	21.7% [17.2 – 27.6]	16.7% [14.3 – 19.8]
PDO	11.3% [4.3 – 18.9]	12.0% [7.3 – 16.3]	11.7% [7.6 – 15.2]	17.7% [14.6 – 20.4]
PDN	3.8% [0.8 – 8.9]	9.9% [5.9 – 14.2]	10.8% [7.0 – 14.4]	12.6% [9.7 – 14.6]
GEO	10.2% [4.3 – 17.7]	12.8% [8.5 – 18.5]	9.6% [6.3 – 13.1]	9.7% [7.2 – 11.5]
GEN	3.6% [0.8 – 7.9]	5.5% [2.6 – 8.7]	3.9% [1.9 – 6.5]	2.7% [1.5 – 3.7]
GDO	14.7% [7.6 – 23.4]	7.3% [3.6 – 10.5]	8.7% [5.5 – 11.9]	10.3% [7.8 – 12.2]
GDN	8.5% [3.9 – 16.0]	2.6% [0.8 – 5.1]	4.6% [2.4 – 7.6]	4.4% [3.1 – 6.1]

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

Types définis en table 4.1.

4.9.6 Effet des prix des carburants

L'analyse *ceteris paribus* précédente reposait sur des comparaisons de systèmes de probabilités lorsque les caractéristiques des agents différaient en valeurs d'indicatrices. Au contraire, l'indice agrégé des prix des carburants étant une variable continue, sa variation permet d'évaluer les changements induits sur les parts de marchés des différentes alternatives en termes d'élasticités. Ainsi, les commentaires suivants relatent l'élasticité des parts de marché lorsque l'on simule une augmentation de 20% du niveau moyen de la variable IPC, aux caractéristiques de l'agent synthétique.

Les résultats d'estimation de la table 4.5 montrent que l'indice global du prix des carburants est un déterminant particulièrement explicatif des niveaux d'utilité relative associés aux véhicules neufs. Notamment, l'augmentation *ceteris paribus* de la variable *IPC* diminue statistiquement la probabilité pour un agent d'acquérir un véhicule EN, quel que soit le niveau de gamme. Au contraire, les voitures PDO présentent un intérêt significativement croissant avec l'augmentation des prix des carburants.

TABLE 4.11 : Effet d'une hausse de 20% de l'indice des prix moyen IPC

Type acquis	Répartition initiale	Répartition finale	Elasticité
PEO	26.1% [24.1 – 30.4]	24.3% [20.7 – 29.0]	-0.35 [-1.25 ; 0.41]
PEN	17.1% [15.0 – 20.8]	12.5% [9.5 – 15.6]	-1.33 [-2.35 ; -0.51]
PDO	16.9% [13.9 – 19.4]	19.2% [15.5 – 23.3]	+0.70 [-0.45 ; 2.27]
PDN	12.4% [9.4 – 14.2]	17.2% [13.8 – 21.8]	+1.93 [0.72 ; 4.57]
GEO	10.1% [7.4 – 11.8]	9.4% [6.2 – 11.9]	-0.34 [-1.89 ; 1.38]
GEN	2.9% [1.6 – 4.0]	1.7% [0.6 – 2.8]	-1.98 [-3.79 ; 0.71]
GDO	10.2% [7.6 – 12.0]	10.8% [7.8 – 13.8]	+0.29 [-1.11 ; 2.41]
GDN	4.3% [3.1 – 6.2]	4.8% [2.9 – 7.2]	+0.47 [-1.82 ; 3.52]

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS. Types définis en table 4.1. Les élasticités mesurent les sensibilités moyennes des probabilités rapportées à une augmentation de 1% de l'indice moyen des prix des carburants. La répartition initiale (finale) correspond au système de probabilité d'acquisition de l'agent synthétique avant (après) l'augmentation de 20% de la moyenne de la variable IPC.

Les parts de marché les plus sensibles et négativement liées à la hausse du niveau des prix des carburants sont les véhicules EN, dont les élasticités estimées pour les véhicules haut et bas de gamme sont respectivement de -1.98 et -1.33. En niveau, les variations induites correspondantes sont respectivement une diminution

de 1.2 et 4.6 points. Plus globalement, la probabilité estimée d'acquérir un véhicule EN passe de 20% à 14.2% après l'augmentation de 20% de l'indice moyen des prix des carburants. Les mesures d'élasticité des probabilités d'acquisition des voitures EO sont d'amplitude moins élevée que les précédentes, et très voisines : -0.35 pour les véhicules de gamme inférieure, et -0.34 pour les voitures haut de gamme. Dans l'ensemble, la part de marché des véhicules EO après une augmentation simulée de 20% du niveau moyen de la variable IPC passe de 36.2% à 33.7%.

Ainsi, l'effet d'une hausse du prix moyen des carburants se traduit systématiquement par une augmentation de l'attrait des acquéreurs pour les véhicules diesel, qui s'alimentent au carburant le moins cher de l'époque (le gazole). Mécaniquement, elle marque au contraire un désintéressement pour les voitures à essence. Plus spécifiquement, la hausse de l'indice des prix augmente la probabilité pour l'agent synthétique de choisir un véhicule PDO (de 16.9% à 19.2%) avec une élasticité mesurée de +0.7. Par ailleurs, l'effet induit sur la part de marché des voitures GDO (de 10.2% à 10.8%) correspond à une élasticité de +0.29.

Mais ce sont surtout les alternatives d'automobiles DN qui rapportent les sensibilités positives au prix des carburants les plus élevées : leur part de marché passe de 16.8% à 22% sous l'impulsion de la hausse de 20% du niveau moyen de la variable IPC, établissant l'élasticité à +1.55. Cette valeur particulièrement élevée tient surtout à l'augmentation induite de la probabilité pour l'agent moyen de choisir une voiture PDN, qui présente une élasticité estimée à +1.93 (contre +0.47 pour celle d'acquérir une automobile GDN). Au niveau agrégé, le modèle estime que l'agent moyen a, après la hausse simulée de \overline{IPC} ¹⁰⁴, près de 1.4 fois plus de chances qu'initialement de choisir un véhicule diesel plutôt qu'essence ($\widehat{OR}_{D-E}^{1.2\overline{IPC}-\overline{IPC}} = 1.39$).

4.10 Conclusion

Dans ce chapitre, nous avons ajusté la version bayésienne du modèle probit multinomial décrite par NOBILE (1998) sur les données françaises de Parc Auto couvrant la période 1997 - 2001, afin de décrire le comportement d'acquisition automobile des résidents français. Les résultats mettent en lumière le rôle prédominant des caractéristiques de l'automobile précédente sur le choix du

¹⁰⁴ \overline{IPC} pour la moyenne de la variable IPC sur l'échantillon.

véhicule, marquant la fidélité significative des agents pour le type de leur ancien modèle. Par ailleurs, l'effet du genre montre une préférence plus élevée des femmes pour les automobiles essence de gamme inférieure. L'influence du revenu est appréhendée au regard de la différence des comportements d'acquisition, toutes choses égales par ailleurs, entre les différentes catégories professionnelles (CSP). Fort logiquement, les résultats *ceteris paribus* mesurent une préférence significative pour les véhicules neufs des agents issus des classes les plus riches (cadres, profession intermédiaires), relativement à ceux appartenant aux classes plus modestes (ouvriers, sans activité professionnelle). La modélisation permet également de dégager un comportement atypique des Franciliens (spécialement les résidents des couronnes), qui vouent une préférence plus marquée que les Provinciaux pour les véhicules essence. Enfin, la progression du niveau des prix des carburants marque un report significatif des choix vers les véhicules diesel au détriment des voitures roulant avec un supercarburant.

Nous acceptons la remarque qu'un modèle probit trivarié aurait pu être appliqué compte tenu de la nature des choix modélisés, ceux-là recoupant trois caractéristiques (gamme, carburant, état), chacune déclinable en deux versions (bas vs haut, supers vs gazole, occasion vs neuf).

En outre, l'acquisition d'une automobile est également déterminée par un ensemble de fonctions que l'utilisateur souhaite lui attribuer. Un prolongement sans doute bénéfique à notre étude serait d'étudier les motifs d'utilisation principaux comme facteurs explicatifs du choix d'acquisition.

Relativement à la littérature existante, notre modèle peut souffrir de ne pas pouvoir mesurer directement la valorisation par les consommateurs des attributs d'alternatives (prix, puissance...) puisqu'ils sont globalement invariants pour tous les acquéreurs de nos données. Dans notre étude, les effets des ces attributs sur les choix d'acquisition sont « aspirés » par les constantes spécifiques aux alternatives. Une solution pour les révéler aurait été de régresser *in fine* ces constantes sur les différences d'attributs avec l'alternative de référence, un peu à la manière de TRAIN (2005).

Conclusion générale

Conclusion générale

Résumé de la méthode et des résultats

Dans la pluralité des choix économiques individuels, notre intérêt s'est penché côté demande, sur l'analyse désagrégée des comportements de motorisation en France. L'automobile a été le cœur de l'analyse qui, tripolaire, a examiné les fondamentaux suivants : usage, quantité, et qualité.

Après un chapitre introductif dont l'objet concernait la présentation de notre source d'information statistique (l'enquête « Parc Auto ») et son traitement pour la construction d'un panel pondéré 1999-2001, nous avons entrepris de modéliser l'intensité d'usage de l'automobile. A cette fin, nous avons appliqué le calque théorique du modèle d'addiction rationnelle de BECKER et al. (1994) sur nos données de panel. Cela nous a conduit à la résolution d'une équation dynamique de la demande annuelle de kilomètres (ou « l'automobilité ») des ménages par la méthode des moments généralisés. Il en est notamment ressorti qu'on ne peut réfuter l'hypothèse d'addiction, au sens microéconomique, à l'usage de l'automobile.

La recherche suivante s'est attachée à expliquer le degré d'équipement des ménages à l'aide d'un modèle probit ordonné bayésien. La spécification latente autorégressive a permis d'asseoir quelques conclusions dynamiques peu fréquentes dans la littérature des modèles catégoriels discrets. Nous avons ainsi dérivé les effets marginaux à court et long terme du changement résidentiel vers les différentes zones de l'Île-de-France, et calculé les élasticités au revenu à court et long terme des probabilités de non, mono et multi-motorisation.

Le dernier volet d'analyse s'est employé à modéliser le choix qualitatif lors de l'acquisition automobile. L'espace des choix croise trois caractéristiques binaires des véhicules : la motorisation (essence ou diesel), la gamme (basse ou haute), l'état à l'acquisition (neuf ou occasion), ce qui constitue un ensemble de huit modalités. Cohérent avec le principe de la maximisation d'utilité, le modèle probit multinomial bayésien a été appliqué aux occurrences d'acquisition entre 1997 et 2001 dans les vagues de l'enquête Parc Auto. Dans la littérature, ce modèle a été très peu appliqué au profit des modèles logit, ce qui constitue une originalité de notre étude. Les caractéristiques de l'utilisateur principal de chaque véhicule, les attributs de la

voiture remplacée, et l'indice des prix des carburants composent nos facteurs explicatifs des choix d'acquisition.

Pour effectuer un survol des principaux résultats, nous proposons d'aborder quelques thèmes inscrits dans le schéma de l'architecture de la thèse, présenté en introduction.

Vitesse d'ajustement des comportements, effet de la motorisation passée

Dans nos travaux de modélisation, nous avons pris un soin tout particulier à révéler le poids des décisions automobiles passées dans les comportements courants. A tous les niveaux, il est apparu que l'habitude est un facteur visqueux qui retient les unités de consommation dans leur état précédent, empêchant ainsi la fluidité d'ajustement des comportements, et justifiant l'intérêt de distinguer les conclusions à différents horizons (court, moyen, et long terme). Nous avons ainsi pu mesurer que les ménages ne considèrent à court terme que 61% de la conséquence totale de long terme induite par un changement (permanent) de caractéristiques, pour ajuster leur demande de kilomètres. Au bout de quatre ans, 99% de l'effet total de long terme est pris en compte.

L'adaptation du taux d'équipement des ménages nous est apparue sensiblement plus lente encore. L'ajustement instantané induit par un changement permanent dans leurs caractéristiques ne correspond qu'à 49% de l'effet total de long terme, celui-ci n'étant assimilé à 99% qu'au bout de 6 ans.

Egalement, les acquéreurs d'automobile qui remplacent un ancien véhicule tendent à reproduire leur choix passé. Cette fidélité se traduit par une préférence statistiquement plus élevée des acheteurs pour les voitures présentant les mêmes attributs de motorisation (essence ou diesel) et de niveau de gamme (bas ou haut) que la précédente.

L'avènement de la voiture dans la seconde moitié du XX^{ème} siècle en France, et des habitudes automobiles bien ancrées font que le public aura sans doute beaucoup de mal à accepter l'idée d'un transfert brutal de sa mobilité vers les modes de transport public. En complément des leviers économiques (prix et fiscalité) pour orienter progressivement les choix individuels vers un usage moins intensif de l'automobile et vers un équipement plus écologique, les solutions pour

réduire rapidement les émissions de gaz à effet de serre à la sortie des pots d'échappement doivent donc être également envisagées du côté de l'offre, déjà très mobilisée dans ce domaine (carburants verts et technologies hybrides aujourd'hui... pile à combustible et l'hydrogène demain)¹⁰⁵.

L'effet du revenu

Dans le second chapitre, nous avons estimé à +0.10 l'élasticité du trafic en automobile des ménages à leur revenu dans le court terme. Dans le long terme, cette mesure atteint +0.16. Bien que significatives, ces valeurs montrent que l'automobile est un bien dont l'usage est très peu sensible aux variations de revenu, toutes autres choses égales par ailleurs, et notamment à nombre d'automobiles constant dans le ménage. Dans le troisième chapitre, nous avons montré que le revenu des ménages constituait un facteur explicatif du niveau de motorisation, mais dont l'influence apparaît assez limitée lorsque l'on examine les élasticités au revenu des proportions de ménages non, mono, et multimotorisés. Cela ne doit pas surprendre puisque d'une part, ces conclusions sont obtenues en maintenant constants les nombres d'adultes et de permis de conduire, et puisque d'autre part, les ménages les plus modestes peuvent recourir pour s'équiper au marché de l'occasion, où l'offre leur est financièrement plus accessible. Ce point nous est confirmé dans notre analyse des choix d'acquisition automobile : nous observons une propension à choisir un véhicule de seconde main nettement et significativement plus élevée chez les individus faisant partie des catégories professionnelles les plus modestes que chez les ressortissants des catégories les plus aisées. Par exemple, nous mesurons dans le quatrième chapitre qu'un ouvrier a, toutes choses égales par ailleurs, 3.7 fois plus de chances qu'un cadre de faire l'acquisition d'une automobile d'occasion plutôt que neuve.

¹⁰⁵ Notons qu'au sujet de la réduction des émissions automobiles, la solution d'encourager le télétravail pour diminuer la fréquence des déplacements domicile-bureau dans les professions qui peuvent « ramener du travail à la maison », tout en restant connecté à l'entreprise (via internet...), n'est pratiquement jamais abordée, alors qu'elle est économique à plus d'un titre (réduction des coûts et du temps de transport, de la congestion automobile et de la fréquentation des transports collectifs aux heures de pointes...). Citons également la solution du co-voiturage, coordonné dans des plans de déplacements de salariés et organisé par les entreprises.

L'effet du prix des carburants

L'analyse de l'effet des prix des carburants a été conduite sur deux niveaux : celui de l'usage automobile tout d'abord, puis celui du choix d'acquisition.

Lorsque les variations du prix courant sont passagères et non anticipées, nous avons mesuré à l'aide du modèle d'addiction rationnelle que l'élasticité du trafic des voitures essence au prix des supercarburants était en moyenne de -0.21 à court terme, contre -0.09 pour l'élasticité de l'usage des voitures diesel au prix du gazole. Lorsque les variations de prix sont permanentes, ces valeurs se situent respectivement à -0.32 et -0.13 dans le court terme, puis à -0.52 et -0.21 dans un horizon de long terme. Notons qu'il s'agit de résultats essentiellement déterminés par les voitures provinciales, qui constituent à elles seules 82% du parc national.

Concernant les choix d'acquisition, nous avons évalué la sensibilité des parts de marché de huit types de véhicules à l'évolution de l'indice des prix des carburants (IPC). Au cours de notre période d'étude (1997-2001), cet indice est caractérisé par un différentiel de prix quasi-constant et toujours favorable au gazole, d'environ 0.29 €, ce qui nous a empêché de séparer les effets-prix des deux principaux carburants. C'est donc sous cette condition implicite que nos résultats conservent une portée statistique valide. Aux caractéristiques de l'individu moyen, nous avons ainsi pu mesurer que l'élasticité de la part des véhicules essence neufs à l'augmentation de l'IPC est estimée à -1.4, alors que celle des véhicules diesel neufs s'est établie à +1.6. En termes d'accroissement des parts de marché, c'est surtout au segment des véhicules diesel neufs en bas de gamme¹⁰⁶ que bénéficie une augmentation de l'IPC, avec une élasticité mesurée à presque +2.

L'effet de la localisation résidentielle (spécificités franciliennes)

Dans les trois fondamentaux de la motorisation examinés dans ce mémoire, nous nous sommes inquiétés de dégager les dissemblances comportementales imputables à la zone résidentielle des agents. A cet effet, nous avons mis en *vis-à-vis* Paris, les couronnes franciliennes et la Province. Il en ressort les conclusions suivantes :

¹⁰⁶ Ce qui explique notamment pourquoi, dans le contexte d'augmentation du prix des carburant à l'époque de nos données, les constructeurs ont principalement diversifié leur offre sur ce segment.

- S'ils acceptent l'égalité entre Province et grande couronne, puis entre petite couronne et Paris, les tests concluent que l'usage automobile est moins intense, toutes choses égales par ailleurs, en petite couronne qu'en grande couronne.

- A mesure que l'on s'éloigne de Paris vers la Province, on observe une décroissance progressive de la sensibilité moyenne au revenu de l'automobilité des ménages.

- Les sensibilités de l'usage des voitures diesel au prix du gazole sont équivalentes en Province et en grande couronne francilienne. Même si elle reste faible, c'est en petite couronne francilienne que cette sensibilité est la plus élevée. Au contraire, les voitures diesel les moins sensibles aux variations de prix du gazole sont parisiennes.

- Concernant la réactivité aux variations du prix des supercarburants, nous avons observé en moyenne que les voitures essence parisiennes étaient les plus sensibles (contrairement aux voitures diesel), suivies des voitures ressortissantes de la petite couronne francilienne. Les automobiles essence les moins sensibles sont provinciales et en grande couronne parisienne.

- L'équipement automobile progresse à mesure que les densités de population et la présence des transports collectifs diminuent. C'est donc en toute logique que nous avons observé, toutes choses égales par ailleurs, qu'un ménage parisien est moins fréquemment motorisé qu'un ménage de petite couronne, lui-même moins motorisé qu'un ménage résidant en grande couronne francilienne. Particulièrement au ménage moyen parisien, la simulation de son transfert vers la petite couronne lui fait perdre dans le long terme 25 points sur sa probabilité d'être non-équipé (64%), 50 points s'il se relocalise en grande couronne. Lorsque l'on simule, à l'inverse, un transfert vers Paris des ménages moyens de petite et de grande couronne, on mesure une forte tendance à la démotorisation, leur probabilité estimée d'être non-équipé passant respectivement de 0.24 à 0.47, et de 0.05 à 0.42. En Île-de-France, seul le transfert d'un ménage moyen depuis la petite vers la grande couronne génère à long terme une probabilité de multi-motorisation supérieure à 0.10.

- Lors de l'acquisition automobile, les résidents de la grande couronne francilienne marquent une préférence significativement moindre que les Provinciaux pour les voitures diesel en bas de gamme, ce qui profite aux voitures essence en bas de gamme. En petite couronne, les acquéreurs vouent une préférence significativement plus forte qu'en Province pour tous les véhicules essence, ce qui s'explique par leur usage moins intensif de l'automobile. Relativement à la Province,

l'acquisition d'un véhicule diesel à Paris est moins examinée sur le segment des véhicules bas de gamme, et davantage sur celui des véhicules haut de gamme.

- Dans l'ensemble des conclusions, le cas spécifique parisien nous semble plutôt intéressant. La remarquable appréciation relative des Parisiens pour les véhicules diesel en haut de gamme traduit peut être aussi un comportement de « grand rouleur », dont les motivations pourraient être de « s'extraire » de l'Île-de-France et accéder aux endroits non desservies par les transports parisiens, en Province. La faible sensibilité des diésélistes parisiens au prix du gazole serait un signe d'une forte dépendance à l'automobile, notamment pour atteindre les espaces verts provinciaux¹⁰⁷, et ainsi s'affranchir de « l'isolement parisien ».

Limites de notre approche et ouvertures

Et les prix des voitures ?

Le prix des automobiles est une variable dont la discrétion au long de nos travaux peut surprendre. Théoriquement, il n'a pas lieu d'intervenir pour décrire l'usage automobile puisqu'il correspond à un coût fixe et perdu permettant d'accéder à l'automobilité.

Cependant, une variable du prix des automobiles aurait été un déterminant économique légitime pour expliquer les niveaux de motorisation. Au cours de notre période d'observation 1999-2001, la courbe de l'indice mensuel des prix en France métropolitaine pour les automobiles¹⁰⁸ nous est apparue très plate, et encore davantage pour leurs moyennes annuelles. Cette stabilité du prix nous a empêché de mesurer son effet réel sur les états de motorisation.

C'est une observation qui s'est maintenue lorsque nous nous sommes intéressés aux prix d'acquisition par type de véhicule dans le quatrième chapitre. Entre 1997 et 2001, ces prix pour chaque segment montrent une bonne stabilité annuelle si bien que leur manque de variabilité n'aurait peut être pas permis dans notre étude d'identifier leurs effets sur les choix d'acquisition. Techniquement, c'est la très forte colinéarité des variables de prix avec les constantes des modèles qui aurait menacé leur identifiabilité, et donc la convergence de nos méthodes statistiques.

¹⁰⁷ Par exemple, pour les week-ends, pour les vacances...

¹⁰⁸ www.insee.fr, séries 063880257, 063880358, 063880459.

Pour pouvoir dégager l'effet des prix des automobiles, il convient de manipuler des données de plus grande variabilité. A cet effet, une période d'étude particulière, caractérisée par des fluctuations de prix inhabituelles ou plus prononcées, peut être choisie. Une autre solution serait d'allonger la plage temporelle des observations afin d'augmenter les situations de prix auxquels sont confrontés les consommateurs, et saisir avec davantage de précision leurs effets dans les choix d'équipement et d'acquisition.

Trois niveaux d'analyse, étudiés en trois étapes

Notre examen de la motorisation sous trois angles particuliers et en trois temps d'analyse a demandé d'utiliser des modèles microéconomiques, des méthodes statistiques, et des modèles économétriques peu répandus dans la recherche en comportements automobiles. A ce titre, chaque étude est en soi une contribution plutôt originale, ce qui fait une force de notre document, mais aussi une faiblesse.

En effet, notre démarche tripolaire de la demande automobile souffre de ne pas mettre suffisamment en lumière l'interconnexion statistique des trois comportements spécifiquement abordés dans ce mémoire.

Les comportements automobiles ne se déterminent probablement pas en trois temps et sur trois plans indépendants uniquement. Dans le cadre de nos préoccupations, il aurait été assurément plus réaliste de considérer un exercice de modélisation unique, traitant à la fois de la quantité, de la qualité, et de l'usage automobile. Sans doute cela aurait-il été possible d'aménager un tel travail avec les données de Parc Auto, mais probablement au prix d'un allègement des propriétés et de la portée de nos modèles pris individuellement. Toutefois, MANNERING et WINSTON (1985), ainsi que VEKEMAN, BERNARD et BOLDUC (2004) ont déjà proposé des modèles sophistiqués pour l'analyse jointe des comportements de choix et d'usage automobile (modèles dits « discret-continu »)¹⁰⁹.

¹⁰⁹ VEKEMAN et al. (2004) proposent en outre une revue de littérature assez récente sur ces modèles.

Et plus généralement...

Nous avons restreint notre travail à la description de trois fondamentaux « triviaux » de la motorisation, qui n'ont pas vocation à pouvoir expliquer l'ensemble des comportements automobiles en France. Par exemple, nous n'avons pas discuté des motifs d'utilisation de la voiture, qui constituent un pôle important de la demande automobile.

Par ailleurs, nous pouvons regretter de n'avoir mis l'automobile en rapport avec les autres modes de déplacement que sous l'éclairage des différences résidentielles. Par exemple, nos travaux ne permettent pas d'étudier les conséquences sur les choix automobiles franciliens d'une modification de l'offre publique, comme par exemple de la grille tarifaire de la RATP ou de la SNCF. De ce point de vue, recadrer l'automobile dans une perspective plus globale des transports apparaît donc profitable, et dans une perspective économique plus générale également. En effet, l'interaction entre la voiture, la localisation, le logement (et les aménités de son environnement), les autres équipements du ménage, et son mode de vie montre que la portée de l'automobile va bien delà de la sphère des transports.

Références Bibliographiques

- ADDA, J. ET R. COOPER (2000): "Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies," *Journal of Political Economy*, 108(4), pp. 778-806.
- AITKEN, A.C. (1935): "On least squares and linear combinations of observations," *Proceedings of the Royal Statistical Society*, 55, pp. 42-48.
- ALBERT, J. et S. CHIB (1993): "Bayesian analysis of binary and polychotomous data," *Journal of the American Statistical Association*, 88, pp. 669-679.
- ARIAS, C. ET T.L. COX (1998): "Maximum Simulated Likelihood: A brief Introduction for Practitioners," University of Wisconsin-Madison, Department of Agricultural and Applied Economics, Staff Paper 421.
- ARMOOGUM, J. I. (2000): "Correction de la non-réponse et de certaines erreurs de mesures dans une enquête par sondage. Application à l'enquête transport et communications 1993-1994," thèse en statistiques, Bruxelles: Université Libre de Bruxelles.
- AULD C. ET P. GROOTENDORST (2004): "An empirical analysis of milk addiction," *Journal of Health Economics*, 23, pp. 1117-1133.
- BALTAGI, B. H. ET J. M. GRIFFIN (2001): "The Econometrics of Rational Addiction: The Case of Cigarettes," *Journal of Business and Economic Statistics*, American Statistical Association, 19(4), pp. 449-454.
- BALTAGI, B. H. ET J. M. GRIFFIN (2002): "Rational addiction to alcohol: panel data analysis of liquor consumption," *Health Economics*, 11, pp. 485-491.
- BATTISTIN, E., R. MINIACI ET G. WEBER (2000): "What do we learn from recall consumption data?," *IFS Working Papers*, W00/10, London: Institute for Fiscal Studies.
- BAUM, C.F., M.E. SCHAFFER ET S. STILLMAN (2003): "Instrumental variables and GMM: Estimation and testing," *Stata Journal*, StataCorp LP, 3(1), pp. 1-31.
- BECKER, G. ET K.M. MURPHY (1988): "A theory of rational addiction," *Journal of Political Economy*, 96, pp.675-700.
- BECKER, G. S., M. GROSSMAN ET K.M. MURPHY (1994): "An empirical analysis of cigarette addiction," *American Economic Review*, 84, pp. 396-418.
- BECKER, G. S. (1996): *Accounting for Tastes*. Cambridge: Harvard University Press.

- BENTZEN, J., T. ERIKSSON ET V. SMITH (1999): "Rational Addiction and Alcohol Consumption: Evidence from the Nordic Countries," *Journal of Consumer Policy*, 22(3), pp. 257-279.
- BERKOVECK, J. (1985): "Forecasting automobile demand using disaggregate choice models," *Transportation Research B*, 19(4), pp. 315-329.
- BERKOVECK J. ET J. RUST (1985): "A nested logit model of automobile holdings for one vehicle households," *Transportation Research B*, 19(4), pp. 275-285.
- BERRI, A., J-L MADRE ET Y. BUSSIERE (2005): "Étalement urbain et croissance de la circulation automobile des ménages en France. Projection aux horizons 2010 et 2020," *Recherche Transports Sécurité*, 89, pp. 305-326.
- BESAG, J. E. (1974): "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, 36, pp. 192-225.
- BLUNDELL, R. W. ET R. J. SMITH (1991): "Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel: une application au comportement d'investissement des entreprises," *Annales d'Economie et de Statistique*, 20-21, pp. 109-123.
- BOLDUC, D. (1999): "A Practical technique to estimate Multinomial Probit Models in Transportation," *Transportation Research B*, 33, pp. 63-79.
- BOUND, J., D. A. JAEGER ET R. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association*, 90, pp. 443-450.
- BROWNSTONE, D., D.S. BUNCH, T.F. GOLOB ET W. REN (1996): "A Transactions Choice Model for Forecasting Demand for Alternative-Fuel Vehicles," University of California at Irvine, Institute of Transportation Studies, Working Paper 96-4.
- BROWNSTONE, D., D.S. BUNCH ET K. TRAIN (2000): "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles," *Transportation Research B*, 34(5), pp. 315-338.
- BREUSCH, T. S. ET A.R. PAGAN (1979): "A simple test for heteroskedasticity and random coefficient variation," *Econometrica*, 47, pp. 1287-1294.
- BROOKS, S.P. ET A. GELMAN (1998): "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, pp. 434-455.

- BROOKS, S.P. et G.O. ROBERTS (1998): "Convergence assessment techniques for Markov chain Monte Carlo," *Statistics and Computing*, 8, pp. 319-335.
- BUTLER, J. S. ET R. MOFFIT (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 50, pp. 761-764.
- CAMERON, S. (1999): "Rational Addiction and the Demand for Cinema," *Applied Economics Letters*, 6(9), pp. 617-620.
- CAMERON, S. (1997): "Are Greek Smokers Rational Addicts?," *Applied Economics Letters*, 4(7), pp. 401-403.
- CAWLEY, J.H. (1999): "Rational Addiction, the Consumption of Calories, and Body Weight," PhD. Thesis, Chicago: University of Chicago.
- CHALOUPKA, F. J. (1991): "Rational addiction behavior and cigarette smoking," *Journal of Political Economy*, 99, pp. 722-742.
- CHANARON, JJ. ET Y. LUNG (1995): *Economie de l'automobile*. Paris (Fr): La Découverte, collection "Repères".
- CHIB, S. (1995): "Inference in panel data models via Gibbs sampling," in: P. Sevestre and L. Matyas (éds.), *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, 2nd revised edition. London: Kluwer Academic Publishers.
- CHOO, S. ET P.L. MOKHTARIAN (2002): "The Relationship Of Vehicle Type Choice To Personality, Lifestyle, Attitudinal, and Demographic Variables," Research Report 02-06, Institute of Transportation Studies, University of California.
- CITEPA (2006): "Emission dans l'air en France (Màj. le 14/04/06)," Paris: Centre Interprofessionnel technique de la pollution atmosphérique.
- COLLET, R. (2004): "Car Ownership in France, A bayesian analysis on panel data," présentation pour le séminaire de coopération franco-japonais « Sakura », 2 juil. 2004, Arcueil (Fr).
- COLLET, R. (2006): "Étude du choix d'acquisition d'automobiles: une application du modèle Probit multinomial bayésien à l'enquête française Parc Auto," papier présenté lors des 23^{èmes} Journées de Microéconomie Appliquée, 2 juin 2006, Nantes (Fr).
- COLLET, R. (2007): "Dynamique du niveau de motorisation automobile des ménages français," papier présenté lors des 24^{èmes} Journées de Microéconomie Appliquée, 1^{er} juin 2007, Fribourg (Ch).

- COOK, R. D., S. WEISBERG (1983): "Diagnostics for heteroscedasticity in regression," *Biometrika*, 70, pp. 1-10.
- COWLES, M. K. ET B.P. CARLIN (1996): "Markov Chain Monte Carlo convergence diagnostics: a comparative review," *Journal of the American Statistical Association*, 91, pp. 883-904.
- CRAGG, J. G. (1971): "Some Statistical Models For Limited Dependent Variables With Application to the Demand For Durable Goods," *Econometrica*, 39(5), pp. 829- 844.
- CUMBY, R.E., J. HUIZINGA ET M. OBSTFELD (1983): "Two-step two-stage least squares estimation in models with rational expectations," *Journal of econometrics*, 21, pp. 333-335.
- DARGAY J. (2004): "The Effect of Prices and Income on Car Travel in the UK," paper presented at the World Conference on Transport Research, July, Istanbul.
- DARGAY, J., L. HIVERT ET D. LEGROS (2006): "An investigation of car ownership in Europe based on the European Community Household Panel," 11th International Conference on Travel Behaviour Research, Kyoto (JP), August.
- DEATON, A., J. MUELLBAUER (1980): *Economics and Consumer Behavior*. Cambridge (UK): Cambridge University Press.
- DEBREU, G. (1960): "Review of R.D. Luce individual choice behavior," *American Economic Review*, 50, pp. 186-188.
- DEVROYE, L. (1986): *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
- DUPUY, G. (1999a): "From the "magic circle" to "automobile dependence": measurements and political implications," *Transport Policy*, 6(1), pp. 1-17.
- DUPUY, G. (1999b): *La dépendance automobile. Symptômes, analyses, diagnostic, traitements*. Paris : Anthropos.
- ELTONY, M.N. (1993): "Transport gasoline demand in Canada," *Journal of Transport Economics and Policy*, 27(2), pp. 193-208.
- FONTAINE, H. ET J.P. HUBERT (1998): "Evolution de l'usage automobile et du risque d'accident," *Recherche Transports Sécurité*, 57, pp.7-20.
- GALLEZ, C. (1994): "Modèles de projection à long terme de la structure du parc et du marché de l'automobile," thèse en sciences économiques, Paris: Université de Paris 1.

- GARDES, F. ET C. STARZEC (2002): "Evidence on Addiction Effects from Households Expenditure Surveys: the Case of the Polish Panel," paper presented at the Econometric Society European Meeting, Venice (It), August.
- GARDES, F., G. DUNCAN, P. GAUBERT ET C. STARZEC (2005): "Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of Food Consumption: The Case of American and Polish Data," *Journal of Business and Economic Statistics*, 23(2), pp. 242-253.
- GARDES, F., R. COLLET ET C. STARZEC (2007): "Estimation of Addiction Models for Transport Expenditure," papier présenté lors des 24^{èmes} Journées de Microéconomie Appliquée, 1^{er} juin 2007, Fribourg (Ch).
- GELFAND, A.E. ET A.F.M. SMITH (1990): "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, pp. 398-409.
- GELMAN, A. ET D.B. RUBIN (1992): "Inference From Iterative Simulation Using Multiple Sequences," (with discussion), *Statistical Science*, 7, 457-472.
- GEMAN, S. et D. GEMAN (1984): "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on pattern analysis and machine intelligence*, 6, 721-741.
- GEWEKE, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," In J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (éds.), *Bayesian Statistics 4*. Oxford (GB): Oxford University Press.
- GEWEKE, J., M. KEANE ET D. RUNKLE (1994): "Alternative Computational Approaches to Statistical Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76(4), pp. 609-632.
- GODFREY, L.G. (1999): "Instrument relevance in multivariate linear models," *Review of Economics and Statistics*, 81(3), pp. 550-552.
- GOODWIN, P.B. ET M.J.H. MOGRIDGE (1981): "Hypothesis for a fully dynamic model of car ownership," *International Journal of Transport Economics*, 8, 313-327.
- GOODWIN, P.B. (1992): "A review of new demand elasticities with special reference to short and long run effects of price changes," *Journal of Transport Economics and Policy*, 26, pp. 155-163.

- GOODWIN, P., J. DARGAY ET M. HANLY (2004): "Elasticities of road traffic and fuel consumption with respect to price and income: a review," *Transport Reviews*, 24(3), pp. 275-292.
- GOODWIN, P. (2002): "Are fuel prices important?," Chapitre dans Chatterjee, G., Chatterjee, K. (éds.), *Transport Lessons from the Fuel Tax Protests of 2000*. Aldershot (UK): Ashgate, pp. 85-94.
- GOURIÉROUX, C. ET A. MONFORT (1994): *Simulation Based Econometric Methods*. Oxford (GB): Oxford University Press, Core lectures series.
- GREENE, W. H. (2000): *Econometric Analysis, 5th edition*. Upper Saddle River (NJ): Prentice-Hall.
- GREENE, D.L. ET P.S. HU (1985): "The Influence of the Price of Gasoline on Vehicle Use in Multi-vehicle Households," *Transportation Research Record*, 988, pp. 19-24.
- GROSSMAN, M., F. J. CHALOUPKA ET I. SIRTALAN (1998), "An empirical analysis of alcohol addiction: results from monitoring the future panels," *Economic Enquiry*, 36, pp. 39-48.
- GROSSMAN, M., F.J. CHALOUPKA ET C.C. BROWN (1999): "The Demand for Cocaine by Young Adults: A Rational Addiction Approach," NBER Working Paper 5713.
- HAJIVASSILIOU, V., D. MCFADDEN ET P. RUUD (1996): "Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and Computational results," *Journal of Econometrics*, 72 (1/2), pp. 85-134.
- HANLY, M. ET J. DARGAY (2000): "Car ownership in Great Britain: Panel data analysis," *Transportation Research Record*, 1718, 83-89.
- HANSEN, L. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50(3), pp. 1029-1054.
- HARRIS T. E. (1956): "The existence of stationary measures for certain Markov processes," In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, 2, pp. 113-124. University of California Press.
- HASTINGS, W.K. (1970): "Monte Carlo sampling methods using Markov chains and thier applications," *Biometrika*, 57, 97.
- HAUSMAN, J. ET D. WISE (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 48(2), pp. 403-426.

- HECKMAN, J.J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), pp. 153-161.
- HECKMAN, J.J. (1981): "Heterogeneity and state dependence," in Sherwin Rosen ed., *Studies In Labour Markets*, 91-139. Chicago: University of Chicago Press.
- HENSHER, D.A., F.W. MILTHORPE ET N.C. SMITH (1990): "The demand for vehicle use in the urban household sector: theory and empirical evidence," *Journal of Transport Economics and Policy*, 24(2), pp.119-137.
- HERAN, F. (2001): "La réduction de la dépendance automobile," *Cahiers Lillois d'Economie et de Sociologie*, 37, pp. 61-86.
- HERPIN, N. ET D. VERGER (2000): *La consommation des Français. Tome 2: Transports, Loisirs et Santé*. Paris: La Découverte, collection "Repères".
- HIVERT, L. (1996a): "Diésélisation et nouveaux diésélistes: les évolutions récentes," Actes, 59, Arcueil (Fr): INRETS, pp. 83-95.
- HIVERT, L. (1996b): "Le comportement des nouveaux diésélistes," Rapport de convention mission effet de serre, Arcueil (Fr): INRETS.
- HIVERT, L. (2001): "Le parc automobile des ménages: Etude en fin d'année (1999 à partir de la source PARC AUTO - SOFRES)," Rapport de convention INRETS-ADEME, Arcueil (Fr).
- HIVERT, L. (2004): "Heterogeneity, Changes in Car Use in France According to Fuel Prices and Types," paper presented at the franco-japanese cooperation meeting "Sakura", Kyoto (JP), December.
- HIVERT, L. ET V. CERRI (2004): "De nouveaux « nouveaux diésélistes », Comment évolue le kilométrage automobile de ceux qui passent au diesel dans la seconde moitié des années 90," Rapport d'expertise pour la DRAST et la MIES, 74 pages.
- HIVERT, L., F. LECOUEY, F. BOURRIOT, ET J.L. MADRE (2006): "Le bilan CO₂ des ménages, le cas de l'Île-de-France et de l'arrondissement de Lille," Convention DAEI/SESP – CEREN et INRETS, n° CO04 00056.
- HOCHERMAN, I., J.N. PRASHKER ET M. BEN-AKIVA (1983): "Estimation and use of dynamic transaction models of automobile ownership," *Transportation Research Record*, 944, pp. 134-141.
- HUBER, P. (1981): *Robust Statistics*. New York: John Wiley and Sons.
- SOFRES (2001): Le Parc Automobile 2000 - Module Ménage. Notes de synthèse rédigées pour l'INRETS, Montrouge (Fr).

- SOFRES (2002): Le Parc Automobile 2001 - Module Ménage. Notes de synthèse rédigées pour l' INRETS, Montrouge (Fr).
- JACKMAN, S. (2000): "Estimation and Inference via Bayesian Simulation: an introduction to Markov Chain Monte Carlo," *American Journal of Political Science*, 44(2), pp. 375-404.
- JOHANSSON, O. ET L. SCHIPPER (1997): "Measuring the long run fuel demand of cars: separate estimations of vehicle stock, mean fuel intensity, and mean annual driving distance," *Journal of Transport Economics and Policy*, 21(3), pp. 277-292.
- KITAMURA, R. et, D. BUNCH (1990): "Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components," in M. Koshi (éd.), *Transportation And Traffic Theory*, pp. 476-496. Amsterdam (NL): Elsevier Science Publishing Co.
- KITAMURA, R., T.F. GOLOB, T. YAMAMOTO ET G. WU (2000): "Accessibility and auto use in a motorized metropolis," Presented at the 79th Transportation Research Board Annual Meeting, Washington, DC.
- LALLA, R., K. RICKERSTEN ET P.H. VALE (2003): "Rational addiction for alcohol: a panel study of Norwegian beer drinkers," Working Paper, Department of Economics and Resource Management, Norwegian University of Life Sciences.
- LAUX, F.L. (2000): "Addiction as a market failure: using rational addiction results to justify tobacco regulation," *Journal of Health Economics*, 19(4), pp. 421-437.
- LAVE, C. ET K. TRAIN (1979): "A Disaggregate Model of Auto-Type Choice," *Transportation Research*, 13(a), pp. 1-9.
- LAWSON, D.M.C. (2004): "Gary Becker and the quest for the theory of everything," Working Paper, University of Notre Dame, Notre Dame.
- LERMAN, S. ET C. MANSKI (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with econometric Applications*. Cambridge: MIT Press.
- LESAGE J.P. (1999): "Applied Econometrics using MATLAB," Manuscript, Dept. of Economics University of Toronto. (Available at: <http://www.spatial-econometrics.com/html/mbook.pdf>)
- MADRE, J.L. (1989): "L'automobile et les ménages: originalité et contrastes," Données sociales Île-de-France, Paris: INSEE.

- MAIRESSE, J. (1998): "Les lois de la production ne sont plus ce qu'elles étaient: une introduction à l'économétrie des panels," *Revue Economique*, 39(1), pp. 225-271.
- MANNERING, F. ET H. MAHMASSANI (1985): "Consumer Valuation of Foreign and Domestic Vehicle Attributes: Econometric Analysis and Implications for Auto Demand," *Transportation Research A*, 19(3), pp. 243-251.
- MANNERING, F. ET C. WINSTON (1985): "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization," *Rand Journal of Economics*, 16(2), pp. 213-236.
- MCCULLOCH R. ET P. ROSSI (1994): "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64, pp.207-240.
- McFADDEN, D.L. (1974): "Conditional logit analysis of qualitative choice behavior", in P. Zarembka (éd.), *Frontiers in econometrics*, New York: Academic Press.
- McFADDEN, D.L. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 57, pp. 995-1026.
- METROPOLIS N., A.W. ROSENBLUTH, M.N. ROSENBLUTH, A.H. TELLER ET E. TELLER (1953): "Equation of state calculation by fast computing machines," *Journal of Chemical Physics*, 21(6), pp. 1087-1092.
- MEURS, H. (1990): "Trip generation models with permanent unobserved effects," *Transportation Research*, 24B, 145-158.
- MOBILIA, P. (1993): "Gambling as a Rational Addiction," *Journal of Gambling Studies*, 9(2), pp. 121-151.
- MOHAMMADIAN, A. ET E.J. MILLER (2003): "Empirical investigation of household vehicle type choice decisions", *Journal of the Transportation Research Record*, 1854, pp. 99-106.
- NADER, R. (1966): *Ces voitures qui tuent*. Paris: Flammarion.
- NEWMAN, P. ET J. KENWORTHY (1989): *Cities and automobile dependence*. Sidney (Au): Gower technical.
- NEWMAN P. (1995): "Can we overcome automobile dependence?," *Cities*, 12(1), pp. 53-65.
- NOBILE, A., C.R. BHAT et E.I. PAS (1997): "A Random Effects Multinomial Probit Model of Car Ownership Choice," in: C. GATSONIS et al. (eds.), *Case Studies in Bayesian Statistics 3*, pp. 419 – 434. New York: Springer.

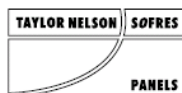
- NOBILE, A. (1998): "A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model," *Statistics and Computing*, 8, pp. 229-242.
- ORFEUIL, J.P. (2000): "L'évolution de la mobilité quotidienne – Comprendre les dynamiques, éclairer les controverses," *Synthèse INRETS*, 37, Arcueil (Fr): INRETS.
- OUM, T.H., W.G. WATERS ET J. YOONG (1992): "Concepts of price elasticities of transport demand and recent empirical estimates: an interpretative survey," *Journal of Transport Economics and Policy*, 26(2), pp. 139-154.
- PAAP R. ET FRANCES P.H. (2000): "A Dynamic Multinomial Probit Model for Brand Choice with Different Long-run and Short-run Effects of Marketing-Mix Variables," *Journal of Applied Econometrics*, 15-6, pp. 717-744.
- PAPON F. ET L. HIVERT (2006): "Adulterous behaviour within the car-owner couple: some analyses from French panel data on car rental and car sharing within households," Communication to the 11th ICTBR, August, Kyoto (JP),
- POWERS, D.A. ET Y. XIE (2000): *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.
- RIPLEY, B.D. (1987): *Stochastic simulation*. New York: Wiley and Sons.
- ROBERT, C. (1996): *Méthodes de Monte Carlo par Chaînes de Markov*. Paris: Economica.
- ROBERT, C.P. ET G. CASELLA (2004): *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- ROSSI, P. E. ET G. M. ALLENBY (2003): "Bayesian Statistics and Marketing," *Marketing Science*, 22, pp. 304-328.
- ROUWENDAL, J. (1996): "An economic analysis of fuel use per kilometre by private cars," *Journal of Transport Economics and Policy*, 30(1), pp. 3-14.
- RUBIN, D.R., ET P.R. ROSENBAUM (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), pp. 41-55.
- SARGAN, J. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica*, 26(3), pp. 393-415.
- SAUVY, A. (1968): *Les quatre roues de la fortune*. Paris: Flammarion.
- SEVESTRE, P. (2003): *Econométrie des Données de Panel*. Paris: Dunod.
- SHEA, J. (1997): "Instrument relevance in multivariate linear models: A simple measure", *Review of Economics and Statistics*, 79(2), pp. 348-352.
- SHERMAN L. ET C.F. MANSKI (1980): "An empirical analysis of household choice among motor vehicles," *Transportation Research A*, 14 (5/6), pp. 349-366.

- SIMONNET, V. ET V. ULRICH (2001): "La formation professionnelle et l'insertion sur le marché du travail: l'efficacité du contrat d'apprentissage," *Economie et statistiques*, 337/338, pp. 81-95.
- SISTO, A. ET R. ZANOLA (2005): "Rationally Addicted to Cinema and Tv? An empirical investigation of Italian consumers," University of Eastern Piedmont, Department of Public Policy and Public Choice (POLIS), Working Papers 46.
- SPIEGELHALTER, D.J., A. THOMAS, N. BEST, ET W. GILKS (1994): "BUGS, Bayesian Inference Using Gibbs Sampling, version 0.30," MRC Biostatistics Unit, University of Cambridge, Technical report.
- STAIGER, D. ET J.H. STOCK (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65(3), pp. 557-586.
- STERN, S. (1992): "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models," *Econometrica*, 60, pp. 943-952.
- STOUT, W. F. (1974): *Almost sure convergence*. New York: Academic Press.
- TANNER, M.A. ET W.H. Wong (1987): "The calculation of posterior distributions by data augmentation," *Journal of American Statistical Association*, 82, pp. 528-540.
- TANNER, M.A. (1991): *Tools for Statistical Inference*. New York: Springer-Verlag.
- TIERNEY L. (1994): "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, pp.1701-1762,
- TIEZZI S. (2003): "Addiction and Smoking Behaviour in Italy," Department of Economics University of Siena, Working Paper 412.
- TOBIN, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, pp. 24-36.
- TRAIN, K. (1986): *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. Cambridge (MA): MIT Press.
- TRAIN, K. ET C. WINSTON (2007): "Vehicle choice behavior and the declining market share of U.S. automakers," *International Economic Review*, à paraître.
- UNCLES, M. D. (1987): "A beta logistic model of mode choice: Goodness of fit and intertemporal dependence," *Transportation Research*, 21B, pp. 195-205.
- VAN OURS, J.C. (1995): "The Price Elasticity of Hard Drugs: The Case of opium in the Dutch East Indies, 1923 - 1938," *Journal of Political Economy*, 103(2), pp. 261 - 279.
- VARIAN, H.R. (1995): *Analyse Microéconomique*. Bruxelles: De Boeck Université.

- VEKEMAN F., J.T. BERNARD ET D. BOLDUC (2004): "Choix de véhicules et demande de kilométrage des ménages québécois," papier présenté au 44^{ième} congrès de la société des sciences économiques, Loews le Concorde, Québec.
- VERBOVEN, F. (2002): "Quantitative Study to Define the Relevant Market in the Passenger Car Sector," Université catholique de Louvain (B), Rapport final pour la commission européenne.
- VILLANI, A. (1992): "Rational Addiction in the Arts," *Ricerche Economica*, 46(1-2), pp. 41-54.
- WHITE, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, pp. 817-838.
- WICKHAM, J. ET ALII (2002): "Scenarios for a sustainable society: Car transport and the sociology of embedded technologie: Final report of Project SceneSusTech," Employment Research Centre, submitted to European Commission DG Research, Dublin.
- WINSTON, C. (1981): "A disaggregate model of the demand for intercity freight transportation," *Econometrica*, 49, 981-1006.

Annexe :

*Formulaire voiture de l'enquête « Parc
Auto » 2001*



16, rue Barbes
92129 Montrouge Cedex
Tél. vert : 08 00 36 90 58

CDF

VEUILLEZ NOUS RENVoyer CE QUESTIONNAIRE
EN MÊME TEMPS QUE LE QUESTIONNAIRE
« VOUS ET LES TRANSPORTS EN COMMUN »
AVANT LE JEUDI 11 FÉVRIER 2002

CE QUESTIONNAIRE EST
À REMPLIR PAR LE CHEF DE FAMILLE :

65 BT 85 - 2001

VOTRE FOYER ET L'AUTOMOBILE

16

Ce questionnaire est à remplir, même si votre foyer ne dispose pas de voiture

① Répondez au tableau suivant pour chacun des individus de 18 ans et plus de votre foyer.

Répondez colonne par colonne	Vous-même	Votre conjoint	Autre personne du foyer	Autre personne du foyer	Autre personne du foyer	Autre personne du foyer
• Prénom.....						
• Age.....						
• Sexe.....	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2
• A un permis de conduire automobile :						
- oui, et conduit régulièrement.....	1	1	1	1	1	1
- oui, et conduit occasionnellement.....	2	2	2	2	2	2
- oui, mais ne conduit pas.....	3	3	3	3	3	3
- non.....	4	4	4	4	4	4
• Nombre de points perdus sur le permis de conduire au cours des 12 derniers mois.....						

② Au cours des 12 derniers mois, c'est-à-dire depuis le 1^{er} janvier 2000, vous-même ou une autre personne de votre foyer a-t-elle loué, en France, une voiture pour motif personnel ? *Ne sont pas pris en compte ici les locations pour motif professionnel, les locations de véhicule utilitaire ainsi que les locations longue durée (c'est-à-dire de plus de 30 jours).*

• OUI..... 1 ➔ Passez à ③

• NON..... 2 ➔ Passez à ③bis

③ Combien de fois au total, au cours des 12 derniers mois, vous-même ou une autre personne du foyer a loué une voiture en France pour motif personnel ? *Précisez le nombre total de locations, puis le nombre de locations de 1 à 4 jours et le nombre de locations de 5 à 30 jours.*

Répondez colonne par colonne	Nombre total de locations	Locations de 1 à 4 jours	Locations de 5 à 30 jours
• 1 fois.....	1	1	1
• 2 fois.....	2	2	2
• 3 - 4 fois.....	3	3	3
• 5 fois et plus.....	4	4	4

③bis Au cours des 12 derniers mois, avez-vous (vous-même ou une autre personne du foyer), loué un véhicule utilitaire pour motif personnel ?

• OUI..... 1

• NON..... 2 ➔ Passez à ⑥

Si vous avez loué une voiture pour motif personnel ou professionnel, passez à ④, sinon passez à ⑥

④ La dernière fois que vous avez loué un véhicule, était-ce pour remplacer votre véhicule provisoirement immobilisé ?

• OUI..... 1

• NON..... 2

⑤ Dans quel contexte avez-vous loué ce véhicule :

- pour les vacances..... 1
- pour le week-end..... 2
- pour un usage régulier..... 3
- pour un déménagement (transport d'objets encombrants)..... 4
- autre..... 5

6 Aujourd'hui, certains organismes proposent de nouvelles formules globales, où tout est compris dans le loyer mensuel (location du véhicule, assurance, entretien, réparation, assistance,...) qui vous permettent de disposer d'un véhicule à l'année avec la possibilité de changement de véhicule tous les deux ou trois ans et un engagement de reprise de votre ancien véhicule.
Cette idée vous paraît-elle :

• tout à fait intéressante... 1 • plutôt intéressante... 2 • plutôt pas intéressante... 3 • pas du tout intéressante... 4

7 Vous, personnellement, seriez-vous intéressé par ce type de location ?

• tout à fait intéressé(e)..... 1
• plutôt intéressé(e) 2
• plutôt pas intéressé(e)..... 3 ➔ **Passez à 10**
• pas du tout intéressé(e) 4

8 Dans le cadre d'un forfait location, vous seriez plutôt intéressé pour louer une voiture :

• tous les jours... 1 • seulement le week-end... 2 • uniquement en vacances... 3 • autre... 4

9 Approximativement, ce forfait location représenterait quel pourcentage de vos déplacements ?

• 100 % de vos déplacements.. 1 • 50 % de vos déplacements.... 3
• 75 % de vos déplacements.... 2 • 25 % de vos déplacements.... 4


10 DE COMBIEN DE VOITURES DISPOSE-T-ON ACTUELLEMENT DANS VOTRE FOYER ?
Comptez toutes les voitures possédées par l'un ou l'autre des membres de votre foyer, même si la voiture ne sert pas actuellement. Comptez également les voitures « de société » dont votre foyer a la libre disposition, les « camping-car », les « motor-home », les « mini-bus » et les fourgonnettes.


• AUCUNE VOITURE ACTUELLEMENT .. 0 ➔ **Passez directement à la question 83 en page 10**
• 1 VOITURE 1
• 2 VOITURES 2
• 3 VOITURES 3
• 4 VOITURES 4

11 De quel type de voiture s'agit-il ?
Répondez aux questions suivantes en utilisant une colonne par voiture. Si plus de 3 voitures, répondez pour les 3 plus récentes.

	a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
• MARQUE de cette voiture.....			
• MODELE (Ex. : 206, Laguna, Saxo, Golf, Fiesta, Corsa, Punto, etc.).....			
• VERSION (Ex. : GTL, TX, GTI, 11TRE, RXE, SX, XR, GRD, Roland Garros, etc.).....			
• SÉRIE SPÉCIALE (Ex. : Junior, Alizé, Tonic, Match, etc.).....			

12 TYPE DE CARROSSERIE :

berline avec coffre 

berline avec hayon 

• Berline avec coffre.....	1	1	1
• Berline avec hayon.....	2	2	2
• Break.....	3	3	3
• Coupé.....	4	4	4
• Cabriolet (décapotable).....	5	5	5
• Monospace - van.....	6	6	6
• Camping-car ou mobil-home.....	7	7	7
• Utilitaire léger (Fourgonnette).....	8	8	8

Pour répondre aux questions 13 à 18, reportez-vous à la carte grise de cette voiture. L'exemple de carte grise figurant sur la page suivante vous aidera à trouver les différentes informations demandées.

13 CODE CARROSSERIE :

• CI.....	01	01	01
• CI2 P.....	02	02	02
• CI4 P.....	03	03	03
• CABR.....	04	04	04
• BREAK.....	05	05	05
• CIALE.....	06	06	06
• CARAVANE.....	07	07	07
• HANDICAP.....	08	08	08
• FOURGON.....	09	09	09
• NON SPEC.....	10	10	10
• Autre (Précisez).....	11	11	11

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
14	GENRE • VP • VU	1 2	1 2	1 2
15	PUISSANCE FISCALE de cette voiture	CV	CV	CV
16	TYPE MINE			
17	NUMÉRO D'IMMATRICULATION : (4 premiers chiffres) (lettres + numéro de département).....			

Exemple de carte grise

Question 17
N° d'immatriculation

N° IMMATRICULATION
2139 QX 94

DATE
17/05/99

DATE DE 1^{re} MISE EN CIRCULATION
31/03/98

Question 18
Année de construction

NOM Prénoms M. MARTIN Cristophe
NOM d'usage

DOMICILE
COMMUNE

Question 14

GENRE VP
MARQUE RENAULT
N° dans la SÉRIE du TYPE
VF1B4030500724788

TYPE
MRE1102BK612

Question 16
Type mine

Question 13
Code carrosserie

Question 15
Puissance fiscale

<p>Continuez de répondre aux questions suivantes dans la même colonne que précédemment</p>	<p>a Voiture principale du foyer</p>	<p>b 2^e voiture</p>	<p>c 3^e voiture</p>
<p>(24) Cette voiture a-t-elle été : Une seule réponse possible</p> <ul style="list-style-type: none"> Achetée à un particulier.....1..... Achetée à un concessionnaire ou agent de la marque.....2..... Achetée à un concessionnaire ou agent d'une autre marque3..... Achetée à un garagiste indépendant.....4..... Achetée à un mandataire.....5..... Elle n'a pas été achetée, elle a été donnée par une personne extérieure au foyer6..... Autre cas (<i>Pécisez</i>)7..... 	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p> <p>.....6.....</p> <p>.....7.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p> <p>.....6.....</p> <p>.....7.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p> <p>.....6.....</p> <p>.....7.....</p>
<p>(25) Cette voiture a-t-elle été introduite dans le foyer, en 2001, par un nouveau membre du foyer ?</p>	<p>Oui ... 1 Non ... 2</p>	<p>Oui ... 1 Non ... 2</p>	<p>Oui ... 1 Non ... 2</p>
<p>(26) A qui appartient cette voiture actuellement ? (Personne ou société inscrite sur la carte grise) <i>Une seule réponse possible</i></p> <ul style="list-style-type: none"> A moi-même.....1..... A mon conjoint.....2..... A une autre personne du foyer.....3..... A une société de leasing - LOA (Location avec Option d'Achat).....4..... A une entreprise privée.....5..... A un service public, une administration.....6..... Autre cas (<i>Pécisez</i>)7..... 	<p>1 } ➔ <i>Passez à (28)</i></p> <p>2 } ➔ <i>Passez à (28)</i></p> <p>3 } ➔ <i>Passez à (28)</i></p> <p>4 ➔ <i>Passez à (27)</i></p> <p>5 } ➔ <i>Passez à (47)</i></p> <p>6 } ➔ <i>Passez à (47)</i></p> <p>7 } ➔ <i>Passez à (47)</i></p>	<p>1 } ➔ <i>Passez à (28)</i></p> <p>2 } ➔ <i>Passez à (28)</i></p> <p>3 } ➔ <i>Passez à (28)</i></p> <p>4 ➔ <i>Passez à (27)</i></p> <p>5 } ➔ <i>Passez à (47)</i></p> <p>6 } ➔ <i>Passez à (47)</i></p> <p>7 } ➔ <i>Passez à (47)</i></p>	<p>1 } ➔ <i>Passez à (28)</i></p> <p>2 } ➔ <i>Passez à (28)</i></p> <p>3 } ➔ <i>Passez à (28)</i></p> <p>4 ➔ <i>Passez à (27)</i></p> <p>5 } ➔ <i>Passez à (47)</i></p> <p>6 } ➔ <i>Passez à (47)</i></p> <p>7 } ➔ <i>Passez à (47)</i></p>
<p>(27) Qui paye les mensualités du leasing - LOA* ? Une seule réponse possible</p> <ul style="list-style-type: none"> Moi-même, mon conjoint ou une autre personne du foyer....1..... Une entreprise privée.....2..... Autre cas (<i>Pécisez</i>)3..... 	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p>
<p>(28) Indiquez le PRIX D'ACHAT TOTAL de cette voiture, livraison, options et remises comprises :</p>	<p>_____ F</p>	<p>_____ F</p>	<p>_____ F</p>
<p>(29) Qui a payé la totalité ou la majeure partie de cette voiture ?</p> <ul style="list-style-type: none"> Vous-même ou votre conjoint.....1..... Un autre membre de votre famille.....2..... Un employeur (ou CE).....3..... Autre cas (<i>Pécisez</i>)4..... 	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>
<p>(30) Cette voiture a été payée, en totalité ou en partie, à l'aide d' : <i>Plusieurs réponses possibles</i></p> <ul style="list-style-type: none"> Un apport personnel (ou au comptant)1..... Un prêt d'un parent, d'un ami.....2..... Un prêt d'un employeur (ou d'un comité d'entreprise).....3..... Un crédit, un prêt ou un leasing - LOA* (auprès d'une banque, d'un organisme de crédit, d'une société de leasing - LOA*, d'une société d'assurance ou d'une mutuelle)4..... Autre (<i>Pécisez</i>)5..... <p><i>Si cette voiture a été payée en totalité ou en partie avec un apport personnel, passez à (31), sinon, passez à (32bis).</i></p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p> <p>.....5.....</p>
<p>(31) Pour le paiement de cette voiture, la somme constituant l'APPORT PERSONNEL, provenait : <i>Plusieurs réponses possibles</i></p> <ul style="list-style-type: none"> D'une épargne constituée sur plusieurs mois.....1..... De la revente d'une précédente voiture.....2..... De la revente d'un autre bien (ex. : maison, Sicav...).....3..... Autre (<i>Pécisez</i>)4..... 	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>	<p>.....1.....</p> <p>.....2.....</p> <p>.....3.....</p> <p>.....4.....</p>

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
32	Indiquez le montant total de cet apport personnel	_____ F	_____ F	_____ F
32bis	Si cette voiture a été payée en totalité ou en partie avec un crédit, un prêt, une LOA*, ou un leasing (auprès d'une banque, d'un organisme de crédit, d'une société de leasing - LOA*, d'une société d'assurance ou d'une mutuelle), passez à 33 , sinon passez à 45			
33	Quel type de crédit, de prêt, de LOA* ou de leasing a été utilisé pour payer cette voiture ? <i>Une seule réponse possible</i> • Un crédit auto (affecté spécifiquement à l'achat de cette voiture) 1 } → Passez à 36 • Un prêt personnel (non affecté spécifiquement à un achat automobile) .. 2 } → Passez à 36 • Un leasing, LOA*, crédit-bail ou bail avec option d'achat 3 } → Passez à 35 • Une location longue durée (LLD) 4 } → Passez à 35 • Une ligne de crédit revolving (réserve d'argent ou crédit permanent) 5 } → Passez à 34			
34	La ligne de crédit revolving (réserve d'argent ou crédit permanent) utilisée pour payer cette voiture, a été ouverte : • Avant l'achat de cette voiture 1 } → Passez à 40 • A l'occasion de l'achat de cette voiture 2 } → Passez à 40			
35	Aviez-vous choisi de recourir à un leasing, LOA*, crédit-bail, ou bail avec option d'achat AVANT d'acheter votre véhicule ou vous l'a-t-on proposé au moment de votre recherche de crédit ? • Avant d'acheter le véhicule 1 • Au moment de la recherche du crédit 2			
36	LORS DE LA RECHERCHE du crédit, du prêt, de la LOA* ou du leasing pour le financement de cette voiture, à qui s'est-on adressé ? <i>Plusieurs réponses possibles</i> • A une seule banque 01 • A un seul organisme de crédit 02 • A une seule société de leasing 03 • A plusieurs banques 04 • A plusieurs organismes de crédit 05 • A plusieurs sociétés de leasing 06 • Au vendeur de la voiture 07 • A son assureur 08 • Autres cas (Précisez) 09			
37	Ce CRÉDIT, ce PRÊT, cette LOA* ou ce LEASING a-t-il été obtenu : <i>Une seule réponse possible</i> • Par l'intermédiaire du concessionnaire (ou du garage) lors de l'achat de la voiture 1 • Par l'intermédiaire d'un organisme d'assurance 2 • Directement auprès de la banque ou de l'organisme qui a accordé ce crédit, ce prêt ou ce leasing 3 / Était-on déjà client de cette banque ou de cet organisme avant l'achat de cette voiture ? Oui ... 1 Non ... 2	Pour répondre aux questions suivantes portant sur le financement obtenu, veuillez vous munir du contrat de votre crédit 1 2 3 Oui ... 1 Non ... 2		
38	INDIQUEZ : • Le montant total emprunté grâce à ce crédit, ce prêt ou ce leasing - LOA* _____ F • Sa durée (en nombre de mois) _____ mois • Le montant des remboursements mensuels _____ F par mois			
39	Le coût de ce crédit, de ce prêt, de cette LOA* ou de ce leasing inclut-il les services suivants ? <i>Ne pas tenir compte des services souscrits en dehors</i> • L'assurance de la voiture 1 • Un contrat d'assistance-dépannage 2 • Un contrat d'entretien de la voiture 3 • Une voiture de remplacement en cas d'immobilisation 4 • Autres services (Précisez) 5 • Le coût de ce crédit, de ce prêt ou de ce leasing n'inclut aucun de ces services 6			

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
40	Indiquez le nom de la banque, de l'organisme de crédit, de la société de LOA*, de leasing, de la société d'assurance ou de la mutuelle qui a accordé ce crédit, ce prêt ou ce leasing <i>Notez en clair et reportez le code de la liste A ou de la liste B ci-jointe selon le cas.</i>	_____	_____	_____
41	Pour quelles raisons a-t-on choisi cette banque, cet organisme de crédit, cette société de LOA*, de leasing, cette société d'assurance ou cette mutuelle ? <i>Plusieurs réponses possibles</i> • Parce qu'on a eu la possibilité de négocier le taux de crédit • Parmi les organismes prospectés, c'est celui qui offrait le taux de crédit le plus bas • Sur les conseils du concessionnaire • C'est l'organisme qui offrait le plus de services • On était déjà client de cette banque ou de cet organisme • Parce que c'était plus simple de traiter tout sur place (achat et financement du véhicule) • Autres (Précisez) 01 02 03 04 05 06 07 01 02 03 04 05 06 07 01 02 03 04 05 06 07
42	Et quelle est la PRINCIPALE RAISON pour laquelle vous avez choisi cette banque, cet organisme de crédit, cette société de LOA*, de leasing, cette société d'assurance ou cette mutuelle ? <i>1 seule réponse possible</i> • Parce qu'on a eu la possibilité de négocier le taux de crédit • Parmi les organismes prospectés, c'est celui qui offrait le taux de crédit le plus bas • Sur les conseils du concessionnaire • C'est l'organisme qui offrait le plus de services • On était déjà client de cette banque ou de cet organisme • Parce que c'était plus simple de traiter tout sur place (achat et financement du véhicule) • Autres (Précisez) 01 02 03 04 05 06 07 01 02 03 04 05 06 07 01 02 03 04 05 06 07
43	Quelle note de 1 à 10 donneriez-vous aux aspects suivants du financement obtenu pour cette voiture ? <i>La note 1 signifiant que vous êtes très insatisfait(e), la note 10 que vous êtes très satisfait(e).</i> • Les conditions du crédit (coût réel, mensualité, durée) • La compétence des interlocuteurs • Le choix des financements proposés
44	Si c'était à refaire, choisiriez-vous de nouveau cette même banque, organisme de crédit, société de LOA*, de leasing, société d'assurance ou mutuelle ? • Oui, certainement • Oui, probablement • Non, probablement pas • Non, certainement pas 1 2 3 4 1 2 3 4 1 2 3 4
45	Disposez-vous actuellement pour cette voiture : • D'un contrat d'extension, d'une prolongation de garantie • D'un contrat de service ou d'entretien	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2
46	Cette voiture est-elle assurée : • au tiers • tous risques 1 2 1 2 1 2
47	Parmi les définitions suivantes, quelle est celle qui correspond le mieux à la situation de cette voiture ? • Avant d'acquérir cette voiture, vous n'aviez pas de voiture dans votre foyer • Cette voiture a été achetée en supplément d'une autre voiture sans pour autant la remplacer • Cette voiture a été achetée pour remplacer une voiture	1 } 2 } 3 }	1 } 2 } 3 }	1 } 2 } 3 }

Passez à 54

Passez à 48

Passez à 54

Passez à 48

Passez à 54

Passez à 48

Continuez de répondre aux questions suivantes dans la même colonne que précédemment	a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
LA PRÉCÉDENTE VOITURE			
(48) Quels étaient : • La marque..... • Le modèle (exemples : 206, Laguna, Saxo, Fiesta...).....
(49) Était-ce : • Un véhicule utilitaire (ex. : fourgonnette, ...)..... • Une voiture DIESEL..... • Une voiture qui roulait au GPL (Gaz de Pétrole Liquéfié).....	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2
(50) Quelle était son année de construction ? C'est-à-dire la date de première mise en circulation.....
(51) Avait-elle été achetée : • neuve • d'occasion..... 1 2 1 2 1 2
(52) En quelle année cette voiture avait-elle été acquise ?
(53) Qu'est devenue cette voiture ? <i>1 seule réponse possible</i> • Vendue à un concessionnaire ou à un agent..... • Vendue à un garagiste non concessionnaire • Vendue à un particulier • Laissée (vendue ou donnée) à une casse..... • Laissée (vendue ou donnée) à l'étranger • Volée • Donnée à un particulier..... • Accidentée (remboursée par l'assurance)..... • Autre cas (<i>Précisez</i>)01.....02.....03.....04.....05.....06.....07.....08.....09.....▼.....01.....02.....03.....04.....05.....06.....07.....08.....09.....▼.....01.....02.....03.....04.....05.....06.....07.....08.....09.....▼.....
LA VOITURE ACTUELLE			
(54) En 2001, quel TYPE DE CARBURANT a-t-on le plus souvent utilisé pour cette voiture ? <i>1 seule réponse possible</i> • Super (97)..... • Super sans plomb (95 ou 98)..... • Gas-oil • GPL (Gaz de Pétrole Liquéfié) / GNV (Gaz Naturel Véhicule) . • Autres (électrique, hybride)	1 → <i>Passez à (58)</i> 2 → <i>Passez à (56)</i> 3 → <i>Passez à (55)</i> 4 → <i>Passez à (54bis)</i> 5 → <i>Passez à (58)</i>	1 → <i>Passez à (58)</i> 2 → <i>Passez à (56)</i> 3 → <i>Passez à (55)</i> 4 → <i>Passez à (54bis)</i> 5 → <i>Passez à (58)</i>	1 → <i>Passez à (58)</i> 2 → <i>Passez à (56)</i> 3 → <i>Passez à (55)</i> 4 → <i>Passez à (54bis)</i> 5 → <i>Passez à (58)</i>
(54bis) Cette voiture est-elle équipée d'un GPL d'origine ou d'un GPL en deuxième monte ? • D'un GPL d'origine • D'un GPL en deuxième monte..... 1 2 1 2 1 2
(55) Cette voiture est-elle un modèle TURBO-DIESEL ? • oui..... • non..... 1 2 <i>Passez à (57)</i> 1 2 <i>Passez à (57)</i> 1 2 <i>Passez à (57)</i>
(56) Quel type de Super sans plomb a-t-on le plus souvent utilisé pour cette voiture en 2001 ? <i>1 seule réponse possible</i> • Super sans plomb 98..... • Super sans plomb 95..... • Les deux..... 1 2 3 1 2 3 1 2 3
(57) Est-elle équipée d'un système à injection directe (TDI, JTD, HDi, dTi, dCi, HPi, ide...) ? • oui..... • non..... 1 2 1 2 1 2
(58) Cette voiture est-elle équipée : • D'un système de climatisation • D'un pot catalytique..... • D'un airbag conducteur • D'un airbag passager..... • D'un système ABS	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		^a Voiture principale du foyer	^b 2 ^e voiture	^c 3 ^e voiture
(59) Et, est-elle équipée : <ul style="list-style-type: none"> • D'un autoradio..... • D'un lecteur de CD..... • D'un changeur de CD • D'un système de navigation (GPS).... 	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2 Oui ... 1 Non ... 2	
<div style="border: 1px solid black; padding: 5px; text-align: center;"> Merci de donner une réponse même approximative aux questions (60), (61), (62) et (63) qui sont pour nous très importantes. </div>				
(60) Quelle est la CONSOMMATION MOYENNE en litres aux 100 kilomètres de cette voiture ?.....	, litres	, litres	, litres	
(61) En moyenne, combien de fois par mois fait-on le plein de carburant de ce véhicule ? Reportez le nombre de fois				
(62) COMBIEN DE KILOMÈTRES A-T-ON PARCOURUS avec cette voiture au cours des 12 derniers mois, c'est-à-dire DEPUIS LE 1^{er} janvier 2001 ? <i>Si cette voiture a été acquise depuis moins d'un an, ajoutez les kilomètres parcourus avec la voiture précédente entre le 1^{er} janvier 2001 et la date d'acquisition de cette nouvelle voiture</i> <ul style="list-style-type: none"> • Cette voiture est en état de marche, mais n'a pas roulé depuis 1 an..... • Cette voiture n'est plus en état de marche depuis 1 an..... 	km 1 2	km 1 2	km 1 2	
(63) Quel est le KILOMÉTRAGE ACTUELLEMENT INSCRIT AU COMPTEUR de cette voiture ? Pour répondre à cette question, regardez le compteur kilométrique de cette voiture. <i>Reportez le kilométrage exact</i>	km	km	km	
(64) Qui est l'UTILISATEUR PRINCIPAL de cette voiture ? <ul style="list-style-type: none"> • Moi-même, le chef de famille ... • Mon conjoint..... • Une autre personne du foyer.... 1 2 3 1 2 3 1 2 3	
(65) Indiquez : <ul style="list-style-type: none"> • Le sexe de l'utilisateur principal • Son prénom • Son année de naissance..... • Sa profession (en clair)..... • L'année d'obtention de son permis de conduire..... 	H ... 1 F ... 2 19 	H ... 1 F ... 2 19 	H ... 1 F ... 2 19 	
(66) L'UTILISATEUR PRINCIPAL de cette voiture reçoit-il une indemnité de transport de son employeur, pour l'utilisation de ce véhicule ? <ul style="list-style-type: none"> • Oui, une somme forfaitaire..... • Oui, sur justificatif des dépenses • Non..... 1 2 3 1 2 3 1 2 3	
(67) Cette voiture est-elle OCCASIONNELLEMENT UTILISÉE PAR D'AUTRES PERSONNES ? <ul style="list-style-type: none"> • OUI • NON 	1 → Passez à (68) 2 → Passez à (70)	1 → Passez à (68) 2 → Passez à (70)	1 → Passez à (68) 2 → Passez à (70)	
(68) Décrivez ci-dessous chacun des utilisateurs occasionnels de cette voiture : <ul style="list-style-type: none"> ■ 1^{er} utilisateur occasionnel : <ul style="list-style-type: none"> • Sexe • Année de naissance..... • Année d'obtention du permis de conduire ■ 2^e utilisateur occasionnel : <ul style="list-style-type: none"> • Sexe • Année de naissance..... • Année d'obtention du permis de conduire 	H ... 1 F ... 2 19 H ... 1 F ... 2 19 	H ... 1 F ... 2 19 H ... 1 F ... 2 19 	H ... 1 F ... 2 19 H ... 1 F ... 2 19 	

Continuez de répondre aux questions suivantes dans la même colonne que précédemment	a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture
69 SUR LE TOTAL DES KILOMÈTRES effectués au cours des douze derniers mois avec cette voiture (et avec la voiture précédente si la voiture actuelle a été acquise depuis moins d'un an), quel a été le POURCENTAGE DE KILOMÈTRES effectués en tout par le ou les UTILISATEURS OCCASIONNELS ? <ul style="list-style-type: none"> • Moins de 10 % • Entre 10 % et moins de 25 % • Entre 25 % et moins de 50 % • 50 % ou plus 1 2 3 4 1 2 3 4 1 2 3 4
70 Généralement (en dehors des périodes de vacances), cette voiture est-elle utilisée : <ul style="list-style-type: none"> • tous les jours • pratiquement tous les jours (4 à 5 jours par semaine) • moins souvent • seulement le week-end • pratiquement jamais 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
71 Cette voiture est-elle utilisée pour les trajets DOMICILE - LIEU DE TRAVAIL (ou lieu d'études) ? <ul style="list-style-type: none"> • OUI • NON 	1 → Passez à 72 2 → Passez à 76	1 → Passez à 72 2 → Passez à 76	1 → Passez à 72 2 → Passez à 76
72 Disposez-vous d'une place de stationnement gratuite sur votre lieu de travail (ou lieu d'études) ? <ul style="list-style-type: none"> • OUI • NON 1 2 1 2 1 2
73 Habituellement, à l'occasion de ces trajets domicile-lieu de travail (ou lieu d'études), cette voiture est-elle utilisée pour : <ul style="list-style-type: none"> • 2 allers et retours par jour • 1 aller et retour par jour • 2 ou 3 allers et retours par semaine • Moins souvent 1 2 3 4 1 2 3 4 1 2 3 4
74 Indiquez le nombre de kilomètres effectués avec cette voiture pour chacun de ces trajets domicile-lieu de travail (ou lieu d'études)..... <i>Comptez 1 aller + 1 retour pour chaque trajet</i>	km	km	km
75 Quel est le temps moyen (en minutes) de ces trajets domicile-travail ? .. <i>Comptez 1 aller + 1 retour pour chaque trajet</i>	mn	mn	mn
76 Cette voiture est-elle utilisée pour : <i>Plusieurs réponses possibles</i> <ul style="list-style-type: none"> • conduire les enfants à l'école, à la crèche • des déplacements professionnels autres que les trajets domicile-travail • aller faire des achats • les loisirs • les sorties, le soir • partir en week-end • partir en vacances 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
77 En 2001, cette voiture a-t-elle été utilisée au moins une fois pour un déplacement professionnel autre qu'un trajet domicile-travail (ex. : visites de fournisseurs, de clients...) ? <ul style="list-style-type: none"> • OUI • NON 	1 → Passez à 78 2 → Passez à 79	1 → Passez à 78 2 → Passez à 79	1 → Passez à 78 2 → Passez à 79
78 Sur le total des kilomètres effectués au cours des douze derniers mois avec cette voiture, quel a été le pourcentage de kilomètres effectués pour des déplacements professionnels ? <ul style="list-style-type: none"> • Moins de 25 % • Entre 25 % et moins de 50 % • Entre 50 % et moins de 75 % • 75 % et plus 1 2 3 4 1 2 3 4 1 2 3 4

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		a Voiture principale du foyer	b 2 ^e voiture	c 3 ^e voiture																														
79	<p>Sur le TOTAL DES KILOMÈTRES effectués depuis le 1^{er} janvier 2001, Y COMPRIS LES KILOMÈTRES EFFECTUÉS AU COURS DES VACANCES, à combien estimez-vous le pourcentage de kilomètres effectués :</p> <ul style="list-style-type: none"> • en ville..... % • sur route (nationales, petites routes de campagne)..... % • sur autoroute..... % • total..... % 	 	 	 																														
80	<p>Pendant la nuit, ce véhicule est-il garé sur un stationnement payant ?</p> <ul style="list-style-type: none"> • OUI..... 1..... • NON..... 2..... 	 	 	 																														
81	<p>Habituellement, où cette voiture est-elle garée la nuit ?</p> <ul style="list-style-type: none"> • dans un garage, un box fermé..... 1..... • sur un emplacement de parking couvert..... 2..... • sur un emplacement de parking en plein air..... 3..... • dans une cour privée, un jardin..... 4..... • dans la rue, sur un parking public..... 5..... 	 	 	 																														
82	<p>Au cours des 5 dernières années, votre foyer s'est-il trouvé au moins 3 mois consécutifs sans aucune voiture à sa disposition ?</p> <ul style="list-style-type: none"> • OUI..... 1 • NON..... 2 																																	
<p align="center">Si votre foyer possède au moins 1 voiture, passez à 86 Si votre foyer ne possède aucune voiture, passez à 83</p>																																		
83	<p>Dans le passé, votre foyer a-t-il disposé d'une voiture ?</p> <p>C'est-à-dire, vous même ou l'un des membres de votre foyer a-t-il déjà disposé d'une voiture, qu'il s'agisse d'une voiture personnelle ou d'une «voiture de société» à libre disposition.</p> <p>En plus des voitures particulières, sont également pris en compte ici les «camping-car», les «motor home», les «mini bus» et les fourgonnettes.</p> <ul style="list-style-type: none"> • OUI..... 1 → Passez à 84 • NON..... 2 → Passez à 86 																																	
84	<p>Depuis quand votre foyer n'a-t-il plus de voiture ?</p> <table border="0"> <tr> <td>Depuis :</td> <td> <ul style="list-style-type: none"> • moins d'un an..... 1 • 1 an à moins de 2 ans..... 2 • 2 ans à moins de 3 ans..... 3 • 3 ans à moins de 5 ans..... 4 </td> <td> <ul style="list-style-type: none"> • 5 ans à moins de 10 ans..... 5 • 10 ans à moins de 20 ans..... 6 • 20 ans ou plus..... 7 </td> </tr> </table>				Depuis :	<ul style="list-style-type: none"> • moins d'un an..... 1 • 1 an à moins de 2 ans..... 2 • 2 ans à moins de 3 ans..... 3 • 3 ans à moins de 5 ans..... 4 	<ul style="list-style-type: none"> • 5 ans à moins de 10 ans..... 5 • 10 ans à moins de 20 ans..... 6 • 20 ans ou plus..... 7 																											
Depuis :	<ul style="list-style-type: none"> • moins d'un an..... 1 • 1 an à moins de 2 ans..... 2 • 2 ans à moins de 3 ans..... 3 • 3 ans à moins de 5 ans..... 4 	<ul style="list-style-type: none"> • 5 ans à moins de 10 ans..... 5 • 10 ans à moins de 20 ans..... 6 • 20 ans ou plus..... 7 																																
85	<p>Pour quelles raisons votre foyer ne dispose-t-il plus de voiture ?</p> <p align="right">Plusieurs réponses possibles</p> <table border="0"> <tr> <td> <ul style="list-style-type: none"> • changement de situation familiale (deuil, divorce,...)..... 01 • déménagement..... 02 • changement de situation professionnelle..... 03 • perte de points sur le permis de conduire..... 04 • difficulté à circuler ou à stationner..... 05 • l'entretien du véhicule revenait trop cher..... 06 </td> <td> <ul style="list-style-type: none"> • pour des raisons de santé..... 07 • préférence pour les deux-roues..... 08 • préférence pour les transports collectifs..... 09 • par crainte de polluer..... 10 • autre (précisez)..... 11 </td> </tr> </table>				<ul style="list-style-type: none"> • changement de situation familiale (deuil, divorce,...)..... 01 • déménagement..... 02 • changement de situation professionnelle..... 03 • perte de points sur le permis de conduire..... 04 • difficulté à circuler ou à stationner..... 05 • l'entretien du véhicule revenait trop cher..... 06 	<ul style="list-style-type: none"> • pour des raisons de santé..... 07 • préférence pour les deux-roues..... 08 • préférence pour les transports collectifs..... 09 • par crainte de polluer..... 10 • autre (précisez)..... 11 																												
<ul style="list-style-type: none"> • changement de situation familiale (deuil, divorce,...)..... 01 • déménagement..... 02 • changement de situation professionnelle..... 03 • perte de points sur le permis de conduire..... 04 • difficulté à circuler ou à stationner..... 05 • l'entretien du véhicule revenait trop cher..... 06 	<ul style="list-style-type: none"> • pour des raisons de santé..... 07 • préférence pour les deux-roues..... 08 • préférence pour les transports collectifs..... 09 • par crainte de polluer..... 10 • autre (précisez)..... 11 																																	
86	<p>Voici une série d'opinions sur l'automobile. Indiquez pour chacune d'entre elles si vous êtes personnellement tout à fait d'accord, plutôt d'accord, plutôt pas d'accord ou pas du tout d'accord.</p> <p align="center">1 seule réponse par ligne</p> <table border="1"> <thead> <tr> <th></th> <th>Tout à fait d'accord</th> <th>Plutôt d'accord</th> <th>Plutôt pas d'accord</th> <th>Pas du tout d'accord</th> </tr> </thead> <tbody> <tr> <td>• La voiture est un simple moyen de déplacement.....</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>• La voiture est un prolongement de votre personnalité.....</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>• La voiture est synonyme d'indépendance et de liberté.....</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>• Être propriétaire d'un véhicule est important.....</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>• Posséder une voiture est une contrainte.....</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> </tbody> </table>					Tout à fait d'accord	Plutôt d'accord	Plutôt pas d'accord	Pas du tout d'accord	• La voiture est un simple moyen de déplacement.....	1	2	3	4	• La voiture est un prolongement de votre personnalité.....	1	2	3	4	• La voiture est synonyme d'indépendance et de liberté.....	1	2	3	4	• Être propriétaire d'un véhicule est important.....	1	2	3	4	• Posséder une voiture est une contrainte.....	1	2	3	4
	Tout à fait d'accord	Plutôt d'accord	Plutôt pas d'accord	Pas du tout d'accord																														
• La voiture est un simple moyen de déplacement.....	1	2	3	4																														
• La voiture est un prolongement de votre personnalité.....	1	2	3	4																														
• La voiture est synonyme d'indépendance et de liberté.....	1	2	3	4																														
• Être propriétaire d'un véhicule est important.....	1	2	3	4																														
• Posséder une voiture est une contrainte.....	1	2	3	4																														

(87) Nous aimerions connaître votre opinion générale pour chacune des marques suivantes.

Si vous avez une opinion très favorable d'une marque, vous donnez la note 10. Si vous avez une opinion défavorable, vous mettez la note 1. Les notes intermédiaires servent à nuancer votre jugement.

1 seule réponse
par ligne

	Tres defavorable 1	2	3	4	5	6	7	8	9	Tres favorable 10
• Audi.....	... 01 02 03 04 05 06 07 08 09 10 ...
• BMW 01 02 03 04 05 06 07 08 09 10 ...
• Citroën 01 02 03 04 05 06 07 08 09 10 ...
• Fiat 01 02 03 04 05 06 07 08 09 10 ...
• Ford 01 02 03 04 05 06 07 08 09 10 ...
• Mercedes..	... 01 02 03 04 05 06 07 08 09 10 ...
• Nissan 01 02 03 04 05 06 07 08 09 10 ...
• Opel.....	... 01 02 03 04 05 06 07 08 09 10 ...
• Peugeot...	... 01 02 03 04 05 06 07 08 09 10 ...
• Renault....	... 01 02 03 04 05 06 07 08 09 10 ...

(88) Pour remplacer ou acquérir votre voiture, quels sont les types de carrosserie que vous pourriez acheter et quels sont ceux que vous n'achèteriez jamais ?

1 seule réponse
par ligne

	Oui, peut-être	Non, jamais
• Une berline 1 2
• Un break..... 1 2
• Un monospace 1 2
• Un coupé cabriolet 1 2
• Un 4 x 4 1 2
• Un utilitaire ou dérivé d'utilitaire 1 2
• Autre 1 2

(89) S'agirait-il d'un véhicule... • Essence... 1 • Diesel... 2 • GPL... 3 • Electrique... 4

(90) Achèteriez-vous ce véhicule... • neuf 1 • d'occasion..... 2

(91) Si vous deviez acheter aujourd'hui une nouvelle voiture, en tenant compte de vos besoins et de vos revenus, quels sont LA marque et LE modèle que vous envisageriez d'acheter ?

• Marque :

• Modèle :

(92) Actuellement, dans votre foyer, combien y a-t-il de personnes qui travaillent ? *Pensez à bien vous compter*

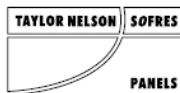
• Personnes • Aucune..... 0

Merci de répondre aux questions suivantes pour chacun des membres de votre foyer.

Repondez colonne par colonne

	Vous-même	Votre conjoint	Autre personne du foyer	Autre personne du foyer	Autre personne du foyer
(93)					
• Prénom.....
• Age.....	<input type="text"/> <input type="text"/> ans	<input type="text"/> <input type="text"/> ans	<input type="text"/> <input type="text"/> ans	<input type="text"/> <input type="text"/> ans	<input type="text"/> <input type="text"/> ans
• Sexe	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2	• H.. 1 • F.. 2
(94) Généralement, en dehors des périodes de vacances, tous les combien cette personne utilise-t-elle les transports en communs ?					
• Tous les jours, y compris le week-end..... 1 1 1 1 1
• Pratiquement tous les jours, hors week-end (c'est-à-dire au moins 4 à 5 fois par semaine)..... 2 2 2 2 2
• En semaine, mais pas tous les jours 3 3 3 3 3
• Seulement le week-end 4 4 4 4 4
• Pratiquement jamais 5 5 5 5 5
• Ne prends jamais les transports en commun 6 6 6 6 6
	Passez à (97)	Passez à (97)	Passez à (97)	Passez à (97)	Passez à (97)

Continuez de répondre aux questions suivantes dans la même colonne que précédemment		Vous-même	Votre conjoint	Autre personne du foyer	Autre personne du foyer	Autre personne du foyer
95	En général, pour les déplacements en transport en commun, combien cette personne dépense-t-elle par mois.....	_____ F	_____ F	_____ F	_____ F	_____ F
96	Utilise-t-elle un titre d'abonnement pour les transports en commun ? • Oui..... • Non.....	1 2	1 2	1 2	1 2	1 2
97	Au cours de l'année 2001, votre foyer a-t-il changé de résidence principale ? • OUI, tout en restant dans la même commune..... 1 • OUI, et a changé de commune..... 2 • NON 3					
98	Etes-vous : • locataire de votre logement 1 • propriétaire de votre logement, avec un crédit immobilier en cours sur ce logement 2 • propriétaire de votre logement, sans crédit en cours sur ce logement 3					
99	Quels sont, parmi les organismes bancaires ou financiers suivants, ceux auprès desquels vous-même ou toute autre personne de votre foyer a un compte-chèques ? <div style="display: flex; justify-content: space-between;"> <div> • BNP 01 • Banque Populaire 02 • Caisse d'Epargne Ecureuil ou Caisse d'Epargne et de Prévoyance 03 • CCP (Caisse d'Epargne de la Poste ou Caisse Nationale d'Epargne)..... 04 • Crédit Agricole..... 05 • Crédit Lyonnais 06 • Crédit Mutuel..... 07 • Société Générale ou SOGENAL 08 • BRA : Banque Régionale de l'Ain 10 • BRO : Banque Régionale de l'Ouest..... 11 </div> <div style="text-align: center; font-size: 2em;">↓</div> <div> • BSD : Banque Scalbert Dupont 12 • BT : Banque Transatlantique 13 • CIAL : Crédit Industriel d'Alsace Lorraine 14 • CICP : Crédit Industriel et Commercial de Paris.... 15 • CIN : Crédit Industriel de Normandie..... 16 • CIO : Crédit Industriel de l'Ouest..... 17 • LB : Lyonnaise de Banque 18 • BLB : Bonnasse Lyonnaise de Banque 19 • SB CIC : Société Bordelaise de CIC 20 • SNVB : Société Nancéenne Varin Bernier..... 21 • Autre..... 09 </div> </div>					
100	Pour nous permettre de classer vos réponses, pouvez-vous indiquer le REVENU ANNUEL NET de votre foyer (EN NOUVEAUX FRANCS) ? <i>Comptez tous les revenus de toutes les personnes de votre foyer (salaires, primes, pensions,...)</i> <div style="display: flex; justify-content: space-between;"> <div> • Moins de 50 000 F 01 • De 50 000 F à moins de 75 000 F 02 • De 75 000 F à moins de 100 000 F 03 • De 100 000 F à moins de 125 000 F 04 • De 125 000 F à moins de 150 000 F 05 • De 150 000 F à moins de 175 000 F 06 • De 175 000 F à moins de 200 000 F 07 </div> <div style="text-align: center; font-size: 2em;">↓</div> <div> • De 200 000 F à moins de 250 000 F 08 • De 250 000 F à moins de 300 000 F 09 • De 300 000 F à moins de 400 000 F 10 • De 400 000 F à moins de 500 000 F 11 • De 500 000 F à moins de 600 000 F 12 • 600 000 F et plus..... 13 </div> </div>					



16, rue Barbès
92129 Montrouge Cedex
Tel. vert : 08 00 36 90 58

LISTE A : LISTE DES BANQUES, ORGANISMES DE CRÉDIT ET SOCIÉTÉS DE LEASING - LOA*

(à utiliser pour (40))

- | | |
|--|---|
| 051 BANQUE DE CRÉDIT GENERAL MOTORS (BCGM) | 075 SOCIÉTÉ GÉNÉRALE |
| 052 BANQUE POPULAIRE | 076 SOVAC |
| 053 BNP (Banque Nationale de Paris) | 077 UNICAR |
| 054 CAISSE D'ÉPARGNE | 078 VAG FINANCEMENT |
| 055 CASDEN BP | 080 CRÉDIT SOCIAL DES FONCTIONNAIRES |
| 056 CCF | 084 CARREFOUR (PASS-SYGMA) |
| 057 CETELEM, AURORE | 085 FRANFINANCE |
| 058 CIC (Crédit Industriel et Commercial) | 086 COFINOGA |
| 059 CGI (Crédit Général Industriel) | 087 CRÉDIT FORD |
| 060 COFICA | 088 MERCEDES FINANCEMENT |
| 061 CRÉDICA VIA - CAVIA | 089 BRA : BANQUE RÉGIONALE DE L'AIN |
| 062 CRÉDIPAR | 090 BRO : BANQUE RÉGIONALE DE L'OUEST |
| 063 CRÉDIT AGRICOLE | 091 BSD : BANQUE SCALBERT DUPONT |
| 064 CRÉDIT DE L'EST | 092 BT : BANQUE TRANSATLANTIQUE |
| 065 CRÉDIT LYONNAIS | 093 CIAL : CRÉDIT INDUSTRIEL D'ALSACE LORRAINE |
| 066 CRÉDIT MUNICIPAL | 094 CICI : CRÉDIT INDUSTRIEL ET COMMERCIAL DE PARIS |
| 067 CRÉDIT MUTUEL | 095 CIN : CRÉDIT INDUSTRIEL DE NORMANDIE |
| 068 CRÉDIT DU NORD | 096 CIO : CRÉDIT INDUSTRIEL DE L'OUEST |
| 069 CRÉDIT UNIVERSEL | 097 LB : LYONNAISE DE BANQUE |
| 070 DIAC / RENAULT BAIL | 098 BLB : BONNASSE LYONNAISE DE BANQUE |
| 071 DIN | 099 SB CIC : SOCIÉTÉ BORDELAISE DE CIC |
| 072 FIAT CRÉDIT | 100 SNVB : SOCIÉTÉ NANCÉENNE VARIN BERNIER |
| 073 SOCRAM | 079 AUTRE. (Précisez dans le questionnaire) |
| 074 SOFINCO | |

LISTE B : SOCIÉTÉS, ORGANISMES D'ASSURANCES et MUTUELLES

(à utiliser pour (40))

- | | |
|--|---|
| 001 ABEILLE ASSURANCES, GROUPE VICTOIRE | 017 GPA |
| 002 ACM (Assurances du Crédit Mutuel) | 018 MAAF |
| 003 AGF (Assurances Générales de France) | 019 MACIF |
| 004 ALLIANZ - VIA ASSURANCES | 020 MAIF |
| 005 AMA (Assurances Mutuelles Agricoles) | 021 MATMUT |
| 006 AXA ASSURANCES, Groupe AXA | 022 MUTUELLES DU MANS - MGF - MMA |
| 007 LE CONTINENT | 023 PACIFICA - PREDICA |
| 008 LA CONCORDE - GENERALI | 024 PFA - PRÉSERVATRICE FONCIÈRE |
| 009 DIRECT ASSURANCES | 025 RHIN ET MOSELLE |
| 010 FILIA MAIF | 027 LA SAUVEGARDE |
| 011 LA FRANCE | 028 UAP |
| 012 Groupe AZUR, GMF | 029 UNI-EUROPE |
| 013 GAN (Groupe des Assurances Nationales) | 030 WINTERTHUR |
| 015 GMF (Garantie Mutuelle des Fonctionnaires) | 031 AUTRE. (Précisez dans le questionnaire) |
| 016 GROUPAMA - SAMDA | |

**NE PAS RENVoyer
À LA SOFRES**

Table des matières

REMERCIEMENTS.....	7
SOMMAIRE	9
INTRODUCTION GENERALE	11
INTRODUCTION GENERALE.....	13
CHAPITRE 1 : LES DONNEES DE L'ENQUETE «PARC AUTOMOBILE».....	25
1.1 CONTEXTE DE L'ENQUETE	27
1.2 STATISTIQUES DESCRIPTIVES	28
1.3 LA MISE EN PANEL DES DONNEES	31
1.3.1 Sélection et biais	31
1.3.2 Redressement par la méthode des scores de propension	33
CHAPITRE 2 : ADDICTION A L'USAGE DE L'AUTOMOBILE ET MESURES D'ELASTICITES	37
2.1 INTRODUCTION	39
2.2 SUPPORTS MICROECONOMIQUES DE LA DEMANDE DE KILOMETRES	41
2.2.1 Le modèle statique de demande de kilomètres	41
2.2.2 Modèles dynamiques de demande : les modèles d'addiction myope et rationnelle de Becker	46
2.2.2.1 Description	46
2.2.2.2 Le modèle d'addiction rationnelle à l'épreuve	52
2.3 MODELES ECONOMETRIQUES ET ELASTICITES.....	55
2.3.1 Elasticités et politique économique.....	55
2.3.2 Mesures du modèle statique.....	55
2.3.3 Mesures du modèle d'addiction myope	56
2.3.4 Mesures du modèle d'addiction rationnelle.....	58
2.4 DONNEES D'APPLICATION ET STATISTIQUES DESCRIPTIVES.....	59
2.4.1 Les données.....	59
2.4.1.1 Traitement des fichiers de véhicules : l'imputation des variables manquantes	60
2.4.2 Statistiques descriptives.....	63
2.5 MODELISATION ET DIFFICULTES	69
2.5.1 Détail du modèle utilisé.....	69
2.5.2 Biais de sélection	71
2.5.3 Hétéroscédasticité	72
2.5.4 Endogénéité	73
2.6 METHODES D'ESTIMATION ET TESTS	73
2.6.1 L'estimateur de la méthode des moments généralisés.....	74
2.6.2 Le test de Breush et Pagan	77
2.6.3 Le test de Hansen	78
2.6.4 Le test de Bound	79
2.7 RESULTATS	80

2.7.1	Qualité des modèles et tests.....	81
2.7.2	Biais de sélection ?.....	82
2.7.3	Addiction et taux de préférence intertemporel	82
2.7.4	Elasticités prix	85
2.7.5	Elasticités au revenu	88
2.7.6	Effet de la localisation des ménages.....	90
2.7.7	Effet du déménagement des ménages.....	90
2.7.8	Effet du vieillissement du parc automobile.....	91
2.7.9	Effet des variables socio-démographiques du ménage	92
2.7.10	Effet de la motorisation / démotorisation des ménages.....	94
2.7.11	Effet des transitions ‘essence ↔ diesel’ lors du remplacement d’un véhicule	99
2.7.12	Limites et critiques de la démarche empirique	100
2.7.13	Comparaisons internationales	101
2.8	CONCLUSION	107

CHAPITRE 3 : DYNAMIQUE DU NIVEAU DE MOTORISATION DES MENAGES

3.1	INTRODUCTION	113
3.2	LE MODELE CATEGORIEL PROBIT ORDONNE DYNAMIQUE	115
3.2.1	Présentation du modèle	116
3.2.2	Décomposition de l’erreur et paramétrage	118
3.2.3	Identification du modèle	119
3.2.4	Traitement des conditions initiales	121
3.2.5	Vraisemblance du modèle	122
3.3	L’INTERPRETATION DU MODELE.....	125
3.3.1	Effets latents et multiplicateurs.....	127
3.3.1.1	Dans le court terme	127
3.3.1.2	Dans le long terme	127
3.3.2	Effets marginaux à court et long terme	128
3.3.2.1	Dans le court terme	128
3.3.2.2	Dans le long terme	129
3.4	ESTIMATION BAYESIENNE MCMC.....	130
3.4.1	Quelques notions bayésiennes d’économétrie	131
3.4.2	Une méthode MCMC : l’échantillonneur de Gibbs.....	133
3.4.2.1	Fonctionnement	133
3.4.2.2	Une extension pratique : l’étape « augmentation data ».....	135
3.4.3	Le diagnostic de convergence de Geweke.....	136
3.4.4	Distributions conditionnelles a posteriori du modèle	137
3.4.4.1	Distributions a posteriori des paramètres	138
3.4.4.2	Distributions a posteriori des variables individuelles	144
3.5	SELECTION DES VARIABLES ET STATISTIQUES DESCRIPTIVES	148
3.6	RESULTATS D’ESTIMATION.....	156
3.7	EXPLOITATION DES RESULTATS	163
3.7.1	Temps d’ajustement des comportements de motorisation	163
3.7.2	L’effet du changement résidentiel	165
3.7.3	Les élasticités au revenu	169
3.8	CONCLUSION	171

CHAPITRE 4 : ETUDE DU CHOIX D'ACQUISITION AUTOMOBILE	175
4.1 INTRODUCTION	177
4.2 REVUE DE LITTERATURE	179
4.3 THEORIE MICROECONOMIQUE DU CHOIX DISCRET	181
4.3.1 Le choix du consommateur	181
4.4 LE MODELE PROBIT MULTINOMIAL (MNP)	183
4.4.1 Formulation de l'utilité aléatoire	184
4.4.2 Identification du modèle MNP	188
4.4.3 Analyse Bayésienne du MNP	190
4.5 LA METHODE MCMC	193
4.5.1 Echantillonnage de Gibbs	194
4.5.2 Les distributions conditionnelles a posteriori	195
4.5.3 Version hybride de l'échantillonnage de Gibbs	200
4.6 CONTROLES DE CONVERGENCE	203
4.6.1 Le diagnostic de GELMAN et RUBIN	204
4.6.2 Le diagnostic de BROOKS et GELMAN	206
4.7 LES DONNEES « PARC AUTO »	206
4.8 RESULTATS D'ESTIMATION	214
4.9 EXPLOITATION DES RESULTATS	219
4.9.1 Le simulateur de fréquence brute CFS	220
4.9.2 Effet de la motorisation passée	222
4.9.3 Effet du sexe de l'acquéreur	224
4.9.4 Effet de la catégorie socio-professionnelle (CSP)	225
4.9.5 Effet de la zone résidentielle	228
4.9.6 Effet des prix des carburants	229
4.10 CONCLUSION	231
CONCLUSION GENERALE.....	235
CONCLUSION GENERALE.....	237
Résumé de la méthode et des résultats	237
Vitesse d'ajustement des comportements, effet de la motorisation passée	238
L'effet du revenu	239
L'effet du prix des carburants	240
L'effet de la localisation résidentielle (spécificités franciliennes)	240
Limites de notre approche et ouvertures	242
Et les prix des voitures ?	242
Trois niveaux d'analyse, étudiés en trois étapes	243
Et plus généralement... ..	244
REFERENCES BIBLIOGRAPHIQUES.....	247
ANNEXE : FORMULAIRE VOITURE DE L'ENQUETE « PARC AUTO »	
2001	259
TABLE DES MATIERES	275
LISTE DES FIGURES	278
LISTE DES TABLES	278

Liste des figures

FIGURE.0.1 : Les nuisances liées au trafic automobile en milieu urbain.....	17
FIGURE 0.2 : Contenu et architecture de la thèse.....	23
FIGURE 2.1 : Détermination de l'output optimal de kilomètres dans la théorie du producteur	46
FIGURE 2.2 : Trois types de consommateurs, trois modèles de consommation.....	51
FIGURE 2.3 : Mesures internationales d'élasticités-prix de la demande de carburant	107
FIGURE 3.1 : Illustration graphique des problèmes d'identification du modèle probit ordonné.....	120
FIGURE 4.1 : Exemple du passage de la distribution a priori à la distribution a posteriori des paramètres dans un modèle probit binaire	192
FIGURE 4.2 : Evolution du prix des carburants en France (1997-2001)	210
FIGURE 4.3 : Traitement multi-chaîne des historiques d'échantillonnage	215

Liste des tables

TABLE 1.1 : Evolution des principaux indicateurs de mobilité et d'équipement automobile.....	29
TABLE 1.2 : Sélection des ménages de la vague 1999 de Parc Auto au panel 1999-2001.....	32
TABLE 2.1 : Résultats des modèles servant l'imputation aux valeurs manquantes/aberrantes.....	62
TABLE 2.2 : Taux de motorisation par année.....	63
TABLE 2.3 : Nombre moyen d'automobiles par ménage par zone.....	63
TABLE 2.4 : Age moyen du parc par type de motorisation	64
TABLE 2.5 : Automobilité des ménages équipés par zone de résidence.....	65
TABLE 2.6 : Répartition dans le parc des types de motorisation par carburant et par zone	66
TABLE 2.7 : Evolution de la consommation de carburant par type de véhicule	66
TABLE 2.8 : Evolution des prix moyens à la pompe des carburants.....	67
TABLE 2.9 : Kilométrage annuel moyen des automobiles par type de motorisation.....	67
TABLE 2.10 : Statistiques descriptives des caractéristiques des ménages	68
TABLE 2.11 : Kilométrages moyens et consommations moyennes des automobiles	69
TABLE 2.12 : Résultats du modèle de sélection	72
TABLE 2.13 : Estimations des modèles statiques et dynamiques de la demande de kilomètres.....	83
TABLE 2.14 : Elasticités au prix du kilomètre de court et long terme	85
TABLE 2.15 : Elasticités au prix du kilomètre par zone de résidence.....	86
TABLE 2.16 : Elasticités au prix des carburants par zone	87
TABLE 2.17 : Elasticités au revenu de court et long terme des modèles	89

TABLE 2.18 : Elasticités au revenu par zone de résidence.....	89
TABLE 2.19 : Effets à long terme d'un changement de zone résidentielle.....	91
TABLE 2.20 : Effets marginaux à court et long terme d'un changement permanent	93
TABLE 2.21 : Pourcentage de l'effet de long terme pris en compte après t périodes.....	95
TABLE 2.22 : Variations à court et long terme de l'automobilité annuelle du ménage par zone de résidence, induites par une automobile neuve additionnelle.....	96
TABLE 2.23 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par une automobile d'occasion additionnelle.....	97
TABLE 2.24 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par la démotorisation partielle d'une automobile de 5 ans	98
TABLE 2.25 : Variations à court et long terme de l'automobilité annuelle des ménages par zone de résidence, induites par la démotorisation partielle d'une automobile de 10 ans	98
TABLE 2.26 : Variation d'automobilité induit par le changement de motorisation d'un véhicule.....	99
TABLE 3.1 : Variables d'intérêt, libellés, et statistiques descriptives.....	149
TABLE 3.1 : Variables d'intérêt, libellés, et statistiques descriptives (suite)	150
TABLE 3.2 : Répartition annuelle des effectifs dans les états de motorisation	150
TABLE 3.3 : Répartition des ménages dans les états de motorisation selon l'activité du chef de famille.....	151
TABLE 3.4 : Nombre d'automobiles dans le ménage selon le nombre de permis de conduire possédés.....	151
TABLE 3.5 : Nombre d'automobiles par zone de résidence.....	152
TABLE 3.6 : Nombre d'automobiles selon la tranche d'âge du chef de famille.....	153
TABLE 3.7 : Nombre d'automobiles par nombre d'adultes dans le ménage.....	153
TABLE 3.8 : Nombre d'automobiles par nombre d'actifs occupés (hors chef de famille) dans le ménage	154
TABLE 3.9 : Nombre d'automobiles par nombre d'enfants mineurs dans le ménage.....	154
TABLE 3.10 : Répartition des effectifs dans les états de motorisation selon le revenu annuel net.....	155
TABLE 3.11 : Transitions des ménages entre états de motorisation 1999-2000 et 2000-2001	156
TABLE 3.12 : Initialisation des paramètres β , β_0 , s_2	157
TABLE 3.13 : Résultats du modèle probit ordonné dynamique	159
TABLE 3.13 : Résultats du modèle probit ordonné dynamique (suite).....	160
TABLE 3.14 : Résultats du modèle probit ordonné de la période initiale	161
TABLE 3.15 : Proportion de l'effet de long terme d'un changement permanent considéré par les ménages	164
TABLE 3.16 : Taux de survie de l'effet initial d'un changement de caractéristiques occasionnel.....	164
TABLE 3.17 : Probabilités de motorisation des ménages synthétiques.....	165
TABLE 3.18 : Effets marginaux de court et long terme d'un déménagement intra Île-de-France	167
TABLE 3.19 : Effets marginaux de court et long terme d'un déménagement vers l'Île-de-France.....	168
TABLE 3.20 : Elasticités au revenu des probabilités d'équipement automobile.....	170

TABLE 4.1 : Codes de désignation des alternatives automobiles	208
TABLE 4.2 : Effectifs automobiles par année d'observation et par type de modalité d'acquisition	208
TABLE 4.3 : Codage et libellé des variables explicatives du modèle MNP	212
TABLE 4.4 : Statistiques descriptives (annuelles) des variables explicatives	213
TABLE 4.5 : Moyennes a posteriori des paramètres β^*	216
TABLE 4.6 : Moyennes a posteriori des paramètres Ω^*	217
TABLE 4.7 : Répartition des probabilités d'acquisition selon le type de véhicule précédent	223
TABLE 4.8 : Répartition des probabilités d'acquisition selon le genre	224
TABLE 4.9 : Répartition des probabilités d'acquisition selon la CSP	226
TABLE 4.10 : Répartition des probabilités d'acquisition selon la zone de résidence	229
TABLE 4.11 : Effet d'une hausse de 20% de l'indice des prix moyen IPC	230